MOLECULAR BASIS OF HETEROSIS IN MAIZE: GENETIC CORRELATION AND 3-DIMENSIONAL NETWORK BETWEEN GENE EXPRESSION AND GRAIN YIELD TRAIT HETEROSIS

A Thesis

by

HUI ZHI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2010

Major Subject: Plant Breeding

MOLECULAR BASIS OF HETEROSIS IN MAIZE: GENETIC CORRELATION AND 3-DIMENSIONAL NETWORK BETWEEN GENE EXPRESSION AND GRAIN YIELD TRAIT HETEROSIS

A Thesis

by

HUI ZHI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Hongbin Zhang
Committee Members,	Joshua Yuan
	C. Wayne Smith
Head of Department,	David D. Baltensperger

December 2010

Major Subject: Plant Breeding

ABSTRACT

Molecular Basis of Heterosis in Maize: Genetic Correlation and 3-Dimensional Network Between Gene Expression and Grain Yield Trait Heterosis.

(December 2010)

Hui Zhi, B.S., Beijing Forestry University

Chair of Advisory Committee: Dr. Hongbin Zhang

Heterosis, or hybrid vigor, refers to the superiority of F_1 hybrid performance over the mean of its parents (mid-parent heterosis) theoretically, or the performance of better parents. It has been discovered in many species of plants and animals as well as in humans, and played an important role in enhanced agricultural production, especially in maize, rice and sorghum although the mechanism have not been elucidated.

We studied the molecular basis of heterosis with a combined genomics and systems biology approach using model organism maize. We profiled the expression of 39 genes that were most differentially expressed (DG) between the mid-parents and their F1 hybrid (Mo17 x B73) in the 13V-satged, developed whole ear shoots of 13 inbred lines and their 22 F1 hybrids grown in the field trails and phenotyped their 13 traits significant for grain yield. The results showed that gene expression varies significantly among inbreds, among hybrids and in heterosis. The gene clustering heat map and gene action networks in inbreds and hybrids were constructed respectively based on their gene expression profile. According to these pattern analyses, we find dramatically difference

between inbreds and their hybrids, although the differential expression varies across different hybrids. Our results also suggest that gene networks are altered from inbreds to hybrids, including their gene contents and wire structures. Last but not least, we have determined the genetic variation correlations between the gene expression and trait performance and constructed the gene networks for the development of 12 of the 13 traits that varied significantly among genotypes. This has led to identification of genes significantly contributing to the performances of the traits, with 1 – 16 genes per trait.

These results have indicated that heterosis results not only from altered expression level of corresponding genes between inbreds and their hybrids, importantly, also from the altered gene action networks and expression patterns. These alternations could be derived from gene actions in a manner of additivity, dominance, over dominance, pseudo-overdominance, epistasis and/or their combinations. Therefore, our findings provide a better understanding of the underlying molecular basis of heterosis. The genes identified for the traits will provide tools for advanced studies of the trait heterosis and could be used as tools for their heterosis breeding in maize. The strategy developed in this study will provide an effective tool for studies of other complicated, quantitative traits in maize and other species.

ACKNOWLEDGEMENTS

I am very grateful to my committee chair, Dr. Hongbin Zhang, who always gives me strong support and valuable suggestions in my master studies. I thank him for giving me the chance to experience this interesting genomic research. I deeply appreciate his great help and advice throughout this research and thesis. Moreover, I would like to acknowledge Dr. Wenwei Xu for kindly providing the plant materials, collecting the RNA tissues and the field phenotypic data, and all suggestions and advice for the research. I would like to thank my committee members, Dr. Joshua Yuan and Dr. C. Wayne Smith, for their guidance and help throughout the course of this research.

My thanks also go to James Huang for help in developing the computer program used in this research, Jialei Xie, for her assistance in statistical analysis of the experimental data, and all colleagues in Dr. Hongbin Zhang's lab for their help and encouragement.

Finally, but not least, I like to express my deep thanks to my parents and grandparents for their endless love, support and encouragements.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Many crop traits important to agriculture, such as yield, quality, hybrid heterosis and resistance to biotic and abiotic stresses, are complicated quantitative traits. They are difficult to study and manipulate due to the fact that they are controlled by multiple genes, readily subjected to environmental variation and can be phenotyped only by quantitative measurements. Although the advent of the DNA marker technology two decades ago has allowed detecting and mapping a number of loci controlling the traits (QTLs), significant limitations exist to further advance the studies, such as the molecular basis of their development and performance. For instance, little is known about what the active expression of a gene in a tissue means with regard to the performance of a quantitative trait and what kind of mechanisms underlies the development of the trait. Such shortage significantly affects the efficient use of quantitative traits for enhanced, long-range crop genetic improvement.

Heterosis is considered as a “miraculous” phenomenon in quantitative traits in agriculture. Introduction of inbred-hybrids to crop production has led to remarkable increase in crop yield over the past 50 years (Duvick, 1999, 2001), especially in maize, rice, and sorghum. It could be predicted that the employment of heterosis in agriculture, particularly its extension in other crop plantsand animal husbands, will continuously play a significant role in enhanced agricultural production. Nevertheless, little is known

---

This thesis follows the style of Plant Physiology.

about its underlying molecular basis, which could be a bottleneck of such efforts.

Heterosis, or hybrid vigor, refers to the phenomenon which describes the superiority of F_1 hybrid performance over the mean of its parents (mid-parent heterosis, MPH) theoretically, or over the performance of better parents (Figure 1). The application of heterosis has revolutionized production of several major crops, including maize, rice and sorghum. For maize, it was estimated that hybrid yield advantage has contributed additional 55 million metric tons (MMT) or approximately 10% of the world total maize production to the world agriculture. The hybrid varieties, in combination with newly developed high-yield inbreds and improved agronomic technologies, have ensured a stable increase in the crop yield. It was estimated that approximately 65% of maize production worldwide are hybrid-based in the late 1990s (Duvick, 1999).

Heterosis has been studied extensively since it was discovered in the early 20^{th} century; however, the molecular basis of heterosis is still obscure. Several hypotheses have been proposed to explain the genetic basis of heterosis, including genome-wide dominance complementation (Davenport, 1908), locus-specific overdominance (ODO) effects (Shull, 1908; East, 1909; Crow, 1952), pseudooverdominace (Jones, 1917), and epistasis (Fisher, 1918;(Goodnight, 1999) (Figure 2). These genetic models have been tested initially through classical quantitative genetic analysis, then quantitative trait locus (QTL) mapping using DNA markers (Stuber et al., 1992; Xiao et al., 1995; Li et al., 2001; Luo et al., 2001; Hua et al., 2003; Lu et al., 2003; Frascaroli et al., 2007; Tang et al., 2010) and recently, genome-wide gene expression profiling (Guo et al., 2006; Huang et al., 2006; Stupar and Springer, 2006; Swanson-Wagner et al., 2006; Guo et al., 2008;

Stupar et al., 2008; Wei et al., 2009). In the dominance hypothesis, the expected phenotype of heterozygotes is between those of the two parents, but deviates from the midpoint of the expected homozygous parents' phenotypes so that it is closer to the expected phenotype of one homozygous parent than the other. In the overdominance hypothesis, the expected phenotype of the heterozygous offspring is outside the range of both homozygous parents. Epistasis hypothesis refers to that hybrid vigor is attributed to the interaction of non-allelic genes from the two parental inbred lines at two or multiple loci in F_1 hybrid. These hypotheses have been further complicated by the phenomenon of pseudo-overdominance, in which hybrid vigor is attributed to the repulsion or *trans*-linkage of recessive alleles with dominant alleles from the two parental inbred lines at multiple loci in F_1 hybrid (Jones, 1917). The genome-wide gene expression analysis showed that all gene action modes, including no dominance (additivity), partial to complete dominance and overdominance, exist in F_1 hybrids relative to their parents, even though only non-additive gene expressions, such as dominance and overdominance, likely contribute to heterosis (Guo et al., 2006; Huang et al., 2006; Stupar and Springer, 2006; Swanson-Wagner et al., 2006; Meyer et al., 2007; Stupar et al., 2008; Wei et al., 2009).

The greatest challenge to decipher the molecular basis of heterosis is likely to causatively link trait heterosis to the molecular events underlying them. Modern QTL mapping has provided a tool to dissect the population average parameters into individual effects residing in different genomic regions or loci and associate the DNA sequences or markers with the trait genetic variation. However, the results from the QTL mapping

experiments are indirect, as they are based on trait variation-marker association and often are inconsistent, from populations to populations and species to species. For example, one study showed dominance complementation as the main contributor to heterosis, whereas others reported overdominance and epistasis as its genetic basis (Xiao et al., 1995; Li et al., 2001; Luo et al., 2001). Stuber et al. (1992) and Lu et al. (2003) suggested that QTL for maize grain yield were almost always associated with heterozygote genotypes, suggesting that overdominance or pseudo-overdominance plays an important role in heterosis. Hua et al. (2003) suggested that single-locus heterotic effects and dominance-by-dominance interactions are the genetic basis of heterosis.

Heterosis has been shown to be an extremely complicated phenomenon in which many genes are likely involved. Therefore, genome-wide approaches, such as global analysis of gene expression using the microarray and serious analysis of gene expression (SAGE), have recently been used to study heterosis in rice and maize. Bao et al. (2005) identified 620 differentially-expressed genes (DGs) in panicles, leaves and roots between a rice superior hybrid (LYP9) and its parents (93-11 and PA64s) by SAGE. Using the same rice hybrid-parent combination, but the microarray technology, Wei et al. (2009) identified 3,540 DGs from seven developmental stages of leaves and panicles. Swanson-Wagner et al. (2006) identified 305 DGs in 14-day seedlings of a maize heterotic F_1 hybrid (Mo17 x B73) and its parental inbreds using a microarray of 13,999 cDNA. Stupar et al. (2006) did a nearly identical microarray study in maize using the same parental inbreds and hybrids and came to a slightly different conclusion that gene expression is mainly additive in heterotic hybrids, with almost no overdominance. Qin et

al. (2010) identified 748 DGs in 13-leaf staged, 40 – 45 mm top ear shoots of a maize heterotic F_1 hybrid (Mo17 x B73) and its inbred midparent, of which 611 genes were not among those previously identified (Swanson-Wagner et al. (2006). However, it remains unknown what the differential gene expression patterns between hybrids and parents means, with regard to phenotypic trait heterosis. To fill in the gap, further research is essential.

Real-time quantitative PCR (RT-qPCR) has been widely used in gene expression analysis, as have been the microarray and SAGE technologies. RT-qPCR amplifies a specific target sequence in a sample and then monitors the amplification progress in a manner of real time using the fluorescent technology. During amplification, the time that the fluorescent signal needs to reach a threshold level (Ct) correlates with the amount of original target sequence, thereby enabling quantification of the sequence in the mRNA population. In comparison, although it can only profile the expression of one gene at a time, RT-qPCR is often used as another technique to validate the results of microarray. Moreover, it has several advantages over the microarray technology for gene expression analysis, such as high sensitivity and high accuracy. It is capable of accurately measuring a specific DNA or RNA molecule in a sample even if there is only a very small quantity. In addition, the final product of RT-qPCR can be further characterized by subjecting it to increasing temperatures to determine when the double-stranded product "melts." This melting point is a unique property dependent on the product length and nucleotide composition (Valasek and Repa, 2005). In this study, we used the RT-qPCR technique to quantify the expression of 39 genes in the developing ear shoots of a

population of 22 hybrids and their 15 inbred parents, thereby determining the relationships between gene expression and phenotypic grain yield trait heterosis. These genes were identified previously in a microarray experiment (Qin et al., 2010) that were most differentially expressed in developing ear shoots between a maize heterotic F_1 hybrid (Mo17 x B73) and its inbred mid-parent.

Maize is economically the most important crop in the USA. The U.S. is the world's largest maize producer and exporter (USDA-FAS, 2007). Furthermore, maize is also the premier choice of experimental organisms for addressing many fundamental biological questions, especially heterosis. First, large numbers of hybrids and large amounts of hybrid seeds can be readily generated, which is required for the research. Second, heterosis has long been used in maize production (Crow, 1998); it is estimated that it contributes to grain yield at 77 kg/ha/year in the U.S. (Duvick, 2005). Third, a wealth of genetic and genomic tools is available in maize. These include inbred lines, recombinant inbred line (RIL) populations, ESTs/unigenes, BAC libraries, integrated physical/genetic maps, genome-wide microarrays, transposon-tagged mutation lines (MaizeGDB) and whole genome sequence (http://www.maizesequence.org/index.html). We took the advantage of these resources and tools to analyze the functions and networks of heterosis candidate genes.

According the central dogma of molecular biology, the transcript (mRNA) level of a gene represents the primary "phenotype" of the gene, while its protein product could be defined the secondary "phenotype" and the visible traits are final phenotypes. Therefore, we could hypothesize that the variation of gene expression is the molecular

basis of the genetic variation of a morphological or physiological trait. The variation of gene expression could be treated as a traditional quantitative phenotypic trait for genetic analysis since it is presented by quantitative measurement, as is a quantitative trait.

The ultimate goals of this study are to determine the genetic correlations in variation between gene expression and grain yield trait heterosis, identify the genes responsible for grain yield trait heterosis and explore the underlying molecular basis of heterosis in maize. In particular, we have accomplished the following research objectives in this study.

Objective 1. To quantify gene expression of 39 genes putatively controlling ear trait heterosis in developing ears of a breeding population consisting of 23 combinations of F_1 hybrid and its inbred parents using RT-qPCR.

Objective 2. To estimate variation of gene expression among different inbred lines, F_1 hybrids and the level of differential expression between F_1 hybrids and their inbred parents.

Objective 3. To determine the correlation in variation between the gene expression heterosis and ear trait heterosis, and identify genes responsible for ear trait heterosis.

Objective 4. To construct the interactive network of the genes responsible for the maize ear trait heterosis

CHAPTER II

MATERIALS AND METHODS

2.1 F_1 hybrid/parental inbred population and trait heterosis phenotyping

A breeding population of F_1 hybrids and their parental inbreds from the maize hybrid breeding program of Texas AgriLife Research, Lubbock, Texas, was used in this study (Table 1). It consisted of 23 F_1 hybrids and 15 parental inbreds. To check their performance, the population was planted at Halfway, 45 miles north of Lubbock, Texas, in 2009. Field plots were 15 feet long, with each plot having 2 rows spaced 30 inches apart. The experiment was replicated two times. The edge of each block was bordered with rows having appropriate inbreeding levels. Plots were machine planted with a precision belt-cone planter that provided relatively even plant spacing. All seeds were treated with fungicide and insecticide to ensure even stands. All cultural practices followed the local practices normally used for hybrid corn production.

During plant growth and development, developing top ear shoots including husks and with a length of 40 – 50 mm were collected randomly at V13 stage. One ear shoot was collected from each replicate of a genotype; therefore, a total of two biologically-replicated ear shoots were collected from the two trial replicates for each genotype. The ear tissues were frozen in liquid nitrogen immediately in the field, transported to laboratory and stored at -80^{o}C before use.

Thirteen traits were phenotyped. These include days to pollens shedding (DTP), plant height in cm (PHT), ear height in cm (EHT), stalk lodging (%) (STL), root lodging

(%) (RTL), husk coverage rating on a 1 to 5 scale (1 = visible tip, 5 = tightly covered) (HUSK), stay green rating on August 22 (SG822), plants with common smut (%) (SMUT), corn earworm feeding damage in cm (CEW), ear length in cm (EL), percentage of molded kernels (MOLD), grain moisture at harvest (MOIST), and grain yield (PLOTYLD).(Table 5)

Mid-parent heterosis (%MPH) of each trait of an inbred and hybrid combination was calculated by the formula:

%MPH (trait A) = {Hybrid (A) / [(Parent$_1$(A)+Parent$_2$ (A))/2]} ×100%

2.2 RNA extraction

The developing top ear shoots collected in the field trial at V13 stage from the maize hybrid breeding population of 23 F_1 hybrids and 15 inbred parents were used for the experiment. As was the phenotyping trial in the field, the ear tissues were sampled from both replicates. Therefore, each genotype was represented by two biological replicates. Total RNA of the ear shoot samples were isolated using the TRIzol reagent (Invitrogen, Carlsbad, CA). Five hundred milligram tissue of each sample was grinded with liquid nitrogen into fine powder and then stored in a 1.5-ml tube at -80^{o}C. One milliliter of the Trizol reagent was added to the tube containing the grinded sample powder and mixed thoroughly. The mixture was stored at room temperature for 15 min and centrifuged at 13,500 rpm, 4^{o}C for 15 min. The supernatant was transferred into a new tube, and 300 μl of chloroform was added to the supernatant and mixed gently. The mixture was kept on ice for 5 min and centrifuged at 13,500 rpm, 4^{o}C for 15 min. The supernatant (about 400 μl) was then transferred into a new tube, an equal amount of

isopropanol (1:1) was added and mixed gently and the mixture was stored at -20^{o}C for 30 – 60 min. Afterwards, the mixture was centrifuged at 13,500 rpm, 4^{o}C for 15 min to pellet total RNA. The supernatant was discarded and two volumes (1.0 ml) of 75% ethanol added to the tube to wash the RNA pellet. The RNA pellet was subjected to centrifugation at 10,000 rpm, 4^{o}C for 10 min. The RNA pellet wash was repeated for two times. The supernatant was discarded, 1 ml of 100% ETOH was added to the tube, and the content was centrifuged at 10,000 rpm, 4^{o}C for 5 min. This step was repeated one more time. Finally, the RNA pellet was air-dried and dissolved in 40 µl DEPC-treated TE.

2.3 Genomic DNA digestion

To remove the DNA potentially contaminated in the RNA, the RNA was treated with DNase I using the DNA-free Kit (Applied Biosystems, Foster City, CA). The treatment was carried out on ice in a 50-µl reaction system containing 5 µl 10 x DNase buffer, 2 µl DNase I (2 U for 20 µg RNA), 40 µl RNA sample and 3 µl H_2O. The reaction was incubated in a 37^{o}C-water bath for 30 min, and then, 10 µl (0.2 volume) of the DNase inactivation reagent was added, mixed thoroughly and incubated at room temperature for 2 min. The reaction was centrifuged at 13,000 rpm, 4^{o}C for 2 min. The supernatant was transferred into a new tube carefully. All RNA samples were diluted to 500 ng/µl after DNase I treatment and stored in -80^{o}C.

2.4. First strand cDNA synthesis

The mRNA in the RNA samples was reverse-transcripted into cDNA by using Oligo dT_{16} primer with the High-capacity cDNA Reverse Transcription Kits (Applied

Biosystems, Foster City, CA). The reverse transcription was conducted in a 40-µl RT-PCR reaction system including 4 µl 10 x RT buffer, 2 µl 25 x dNTP (25 mM), 1 µl Oligo dT_{16} primer (50 µM), 1 µl M-MLV Rase (200 U/µl), 1 µl RNase inhibitor, 1 µl $MgCl_2$ (50 mM), 10 µl H_2O and 20 µl total RNA. The RT-PCR reaction condition was 25^{o}C for 10 min, 37^{o}C for 120 min, 85^{o}C for 5 min and 4^{o}C indefinitely. cDNA samples were then diluted to 0.02 µg/ µl original RNA. cDNA samples were stored in -20^{o}C before use.

2.5 Gene expression analysis using RT-qPCR

Thirty-nine genes that were most differentially expressed or with the highest LogRatios between a maize super heterotic F1 hybrid (Mo17 x B73) and its parental inbreds were analyzed in this study. They were identified using the microarray technology in our previous study (Qin et al. 2010) (Table 2). For each gene, primers for RT-qPCR were designed to produce products in a size range from 75 - 150 bp, to have a primer T_ms of ~60^{o}C, and a primer length ranging from 20 to 24 bp using the software Beacon Designer v7.21 (PREMIER Biosoft International, Palo Alto, CA)(Table 3 and 4). All primers were checked for their uniqueness against a database of the available maize sequences. The forward and reverse primers were diluted into 0.8 µM as their stock solution (20 x). The final concentration of a primer in the RT-qPCR was 0.04 µM. Their PCR products were checked by melting curve analysis and gel electrophoresis (3.5% agarose gel, 40 V for 4 h) using the cDNA templates of Mo17 x B73 and its parental inbreds. The primers that yielded single amplicons with expected sizes were selected for RT-qPCR using the entire inbred and F_1 hybrid population. Each gene was assayed in

replicate for each RNA sample. The genes used as internal controls were selected from β-Actin *(ACT1)*, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH)*, elongation factor 1-alpha (*EF1α)* and high mobility group protein (*HMG)* (Table 3) using the geNorm software (http://medgen.ugent.be/~jvdesomp/genorm/). Those with an *M* value less than 1.5 were selected. cDNA samples prepared above were diluted to 1/10 of its original concentration for RT-qPCR reactions. A 30-µl system for 2 reactions was used for RT-qPCR, including 15 µl MasterMix (Applied Biosystems, Foster City, CA), 1.5 µl 0.8 µM forward primer, 1.5 µl 0.8 µM reverse primer and 12 µl cDNA template so that errors resulted from sample pipetting could be minimized. The reaction condition for RT-qPCR is described in Figure 3, with Step 4 for dissociation curve. Gene expression data was collected and analyzed by the software SDS v2.3 (Applied Biosystems, Foster City, CA). The gene expression level of sample 'A' comparing to the reference gene was calculated by the formula

$2^{-\Delta T} = 2^{-[(\text{Ct gene of interest} - \text{Ct internal control}) \text{ sample A}]}$

(Schmittgen and Livak, 2008)

2.6 Estimation of variation in gene expression among inbred lines, F_1 hybrids and the levels of differential expression between F_1 hybrids and their inbred parents (eMPH)

Thirty-nine genes with same parameters for RT-qPCR amplification and high-quality, consistent RT-qPCR products were selected to estimate the variation of gene expression among inbred lines, among F_1 hybrids and the levels of differential expression between F_1 hybrids and their inbred parents (eMPH). Since two biological

replicates and two technical replicates were applied to each genotype, the data set was fitted to the One-way ANOVA model, with the genotypes considered as the fixed effect. The expression level of each gene was normalized against that of the internal control gene *HMG*, which was the dependent variable, and transformed into $\log_{10}$ data. The JMP version 8.0 (http://www.jmp.com/) was used to perform the ANOVA. The *F* ratio of the ANOVA with a *P*-value ≤ 0.05 was considered as significant variation. To estimate the level of differential expression (eMPH) between F_1 hybrids and their inbred parents, the eMPH was calculated by the formula:

eMPH (expression) = $\{2^{-\Delta T(hybrid)}/[(2^{-\Delta T(parent1)}+2^{-\Delta T(parent2)})/2]\}$ x 100%

2.7 Hierarchical clustering heat map construction

To infer the relationships in gene expression between genotypes and between genes, the mean values of gene expression data of the hybrids orinbreds calculated by the formula $2^{-\Delta T}$ and that of MPH calculated by the formula eMPH (expression) described above were imported into the Multi-experiment Viewer (MeV) version 4.6.1 (www.tm4.org). Hierarchical clustering heat maps were generated with genes and genotype order optimized. Spearman rank correlation with *P*-value ≤ 0.05 was applied for distance metric selection.

2.8 Estimation of variation correlation in expression between 39 genes and between the gene expression and grain yieldtrait phenotypes in hybrids

To determine the expression correlation coefficients pairwisely between the genes studied, a multivariate test was performed between the 39 genes using the JMP version 8.0. A spearman correlation table or a matrix of correlation coefficients that

summarizes the strength of the linear relationships between each pair of response (Y) variables was acquired. Pairwise correlation with coefficients value larger than 0.50 and *P*-value ≤ 0.05 was considered to be related significantly.

The field phenotypic data collected from the inbred and F_1 hybrid population were used for the correlation analysis. Following traits were used in the analysis (see above): days to pollens shedding (DTP), plant height (PHT), ear height (EHT), stalk lodging (%) (STALKL), root lodging (%) (RTL), husk coverage rating on a 1 to 5 scale with 1 forvisible tip and 5 for tightly covered (HUSK), stay green rating on August 22 (SG822), plants with common smut (%) (SMUT), corn earworm feeding damage in cm (CEW), ear length in cm (EL), percentage of molded kernels (MOLD), grain moisture at harvest (MOIST), andgrain yield (PLOTYLD). Only the traits that varied significantly and only the genes whose expressions varied significantly in the ANOVA were participated in the analysis. The variation of a trait was compared against the expression level of every gene to seek for the correlation between them. Only in the case of the correlation with a coefficient ≥ 0.50 and a significance level of $P \leq 0.05$, the gene was considered to be responsible for the trait. If the correlation had a coefficient < 0.50, but was significant at $P \leq 0.05$, the gene was considered to be involved in the trait development.

2.9 Construction of the interactive network of the genes responsible for the maize trait in hybrids

The correlation coefficients were used for the construction of 3-D gene networks by the computer program GeneNet, modified for our research purpose based on the

BioLayout Express Version 3.0 (http://www.biolayout.org/). The higher the correlation coefficients, the more related the two genes or the gene and a trait, while the smaller the correlation coefficients, the farther the relationships between the two genes or the gene and trait. Absolute values of negative coefficients were also used in the construction of the network; negative correlated genes were indicated as red spheres in the network while positive correlated genes were indicated as green spheres. Gene networks of inbred lines, hybrids, and eMPH of gene expression with a threshold of correlation coefficients at 0.6, 0.7, 0.8 and 0.9 were constructed, respectively. Furthermore, the gene expression-trait variation correlations were also included in this analysis and a 3-D network for the trait development was obtained. Traits were indicated as different colors from those of genes in the network. Trait-correlated gene networks of hybrids with a threshold of correlation coefficient ≥ 0.3, 0.4 and 0.5 were constructed, respectively.

CHAPTER III

RESULTS

3.1 Variation of gene expression among inbred lines

To test the hypothesis that genetic variation of gene expression exists among different maize inbred lines, expressions of 39 genes in 13 inbred lines were analyzed by RT-qPCR. The gene *HMG* (X58282.1) was selected as the internal reference and the gene *ACT1* (GQ339773) was selected as the negative control using GeNorm version 3.5 with an M value of 1.367. Since two biological replicates and two technical replicates are applied for each genotype, the data set fitted the One-way ANOVA model, with the genotypes considered as the fixed effect. The expression level of each gene normalized using the expression level of *HMG*, as dependent variable, was treated by Log_{10} transformation. The test was conducted by the JMP version 8.0. The result showed that the negative control *ACT1* did not have significant expression variation among the inbreds (*F* ratio = 1.5348, *P*-value = 0.1473) or hybrids (*F* ratio = 1.4042, *P*-value = 0.1526). In comparison, all of 39 genes studied showed significant expression variation among the 13 inbreds with a *P*-value $\leq$ 0.05, 0.01 or 0.001. Figure 4 shows the distribution of gene expression among the 13 inbreds with *F*-ratios and *P*-values.

3.2 Variation of gene expression among F_1 hybrids

Expression variation of the 39 genes was also tested in the F_1 hybrids of the 13 inbred lines using the JMP version 8.0. Twenty-two F1 hybrids were analyzed, of which 20 were selected by removing outliers. Since two biological replicates and two technical

replicates were applied for each F_1 hybrid, the One-way ANOVA model was used for the data analysis, with the hybrids as the fixed effect. The expression level of each gene was normalized by using that of the *HMG* gene and was transformed by Log_{10} transformation. Similarly, no significant variation was detected for the expression of the negative control gene *ACT1* whereas all of 39 genes studied showed significant expression variation among the 20 hybrids, with a *P*-value $\leq$ 0.05, 0.01 or 0.001. Figure 5 shows the distribution of the gene expression variation among the 20 hybrids with *F*-ratios and *P*-values.

3.3 Variation of eMPH among Inbred/F1 Hybrid Combinations

Gene expression MPH of 39 genes among the 20 combinations that differed significantly in gene expression was analyzed to estimate the variations of the gene expression in F_1 hybrids relative to their inbreds. The expression MPH was calculated by the formula "eMPH" (see Materials and Methods). The data of eMPH was transformed in Log_{10} data to fit One-way ANOVA model. No significant variation was detected for the eMPH of the negative control gene *ACT1*. Twenty-nine of the 39 genes studied showed significant variation in eMPH, with a *P*-value $\leq$ 0.05, 0.01 or 0.001. Figure 6 shows the distribution of eMPH variation of each gene among the 20 inbred/hybrid combinations.

In addition, the variation of eMPH was shown to be more conservative than the gene expression variation among the inbred lines or hybrids only. Of the 39 genes studied, 10 (25.64%) did not show significant variation in eMPH among the 20 inbred/hybrid combinations. The 10 genes that did not show significant variation were

3U9 (hypothetical protein F16F4.11), 1U5 (NA), 1U7 (NA), 1U9 (NA), 2D1 (glucose-1-phosphate adenylyltransferase), 2U3 (drought-induced hydrophobic protein), 2U4 (catalase isozyme 3), 2U9 (ESTs AU083541), 3D2 (At1g69640/F24J1.22), and 3U10 (putative branched chain alpha-keto acid dehydrogenase E2 subunit).

3.4 Cluster analysis of inbred lines, hybrids and genes based on gene expression patterns

The 39 genes studied were clustered based on their expression patterns among 13 inbred lines or among 20 hybrids and the inbreds or hybrids clustered based on the expression patterns of the 39 genes by constructing gene expression heat maps. Spearman correlations were used for the clustering, with a *P*-value ≤ 0.05, 0.01 or 0.001. The genes that had closer expression patterns among genotypes, or the genotypes that had closer expression patterns among the 39 genes were clustered into a single group and those that had different expression patterns were clustered into different clades in the analysis. This was confirmed by the fact that the F_1 hybrids sharing a parent were clustered together.

Figure 7 shows the relationships between the 13 inbreds and Figure 8 shows the relationships between the 22 hybrids based on the expression patterns of the 39 genes. The inbreds were clustered into two large groups or clades, with one containing five inbreds and the other containing 7 inbreds. As a positive control, the inbreds B73 and Mo17 that yields super heterotic hybrids when crossed were clustered into different clusters. The 22 hybrids were also clustered into two large groups or clades, with each having 11 hybrids. As expected, the F_1 hybrids sharing a parent tended to be clustered

together, suggesting their closer correlation in genome constitution. For instance, Hybrids "ARN0902×S2B73" and "ARN0902×CUBA-1" shared the female parent ARN0902 and were clustered into a close group. Nevertheless, when a comparison was made between Figures 7 and 8, the F_1 hybrids, derived from the crosses between the parental inbreds clustered into the same large group might be clustered into different large groups. The opposite was also found to be true. These results suggest the expression pattern of a gene may be significantly altered after they were brought into a new genetic environment by crossing.

To further confirm this conclusion, we also conducted the cluster analysis of the hybrids and their parental inbreds based on the expression patterns of the 39 genes (Figure 9). The result was unexpected; all 22 F_1 hybrids were separated from their parent inbreds. For instance, hybrid "CUBA-1×B110" was far away from its parental inbreds, "CUBA-1" or "B110", but close to "CUBA-1×Tx205" and "CUBA-1×BR-1". This further confirmed the variation of the gene expression level and patterns from inbreds to hybrids. On the other hand, most of the genes studied showed different expression patterns between inbreds and hybrids. The range of gene expression level of 39 genes in inbreds was from 0 to 81.79, while that in hybrids was from 0 to 179.17; it was nearly two-fold larger in hybrids than inbreds. The exceptions were detected on for four of the 39 genes, including 2D6 (nucleasome/chromatin assembly factor D protein), 2D9 (unknown protein), 2D1 (glucose-1-phosphate adenylyltransferase), and 1D9 (proliferating cell nuclear antigen). These genes showed similar expression level among inbreds and hybrids.

3.5 Cluster analysis of the hybrids and their "mid- parents"

The 22 hybrids were also clustered with their parents using the expression levels of the 39 genes in the hybrids and the means of their expression levels in the two parents or "parent 1 + parent 2 artificial hybrids" (Figure 10). It was also unexpected that all 22 F_1 hybrids, except for CUBA-1 x Tx205, were clustered into different groups from their "artificial hybrids". For instance, hybrid "B73×Mo17" was clustered into different groups from "M-B73+Mo17". In comparison, the artificial hybrids sharing a parent tended to be clustered together, suggesting the variation of the gene expression patterns from parents to hybrids. In addition, most of the genes showed different expression levels between hybrids and their "artificial hybrids".

3.6 Cluster analysis of the hybrids and genes based on eMPH

Hybrids and genes were categorized by hierarchical clustering based on eMPH (Fig 11). Hybrid B73×Mo17 had the most genes, 23 out of the 39 genes (58.97%), that were up-regulated, followed by DK-5 x C273A63C-1A, whereas the hybrid DK-7 x S2B73-1 had the least number of genes, 1 of the 39 genes, that was up-regulated. The hybrids sharing a parent tended to be clustered together, showing similar patterns and levels of expression variation from parents to hybrids. Of the 39 genes, 3D5 (putative enoyl-ACP reductase) and 3D7 (putative diphosphate-fructose-6-phosphate 1-phosphotransferase alpha chain) were both up-regulated in 17 out of the 21 hybrids (80.95%) whereas 2D6 and 2D1 were both down-regulated in 20 out of the 21 hybrids (95.24%).

3.7 Pairwise correlation and gene network of gene expression in inbreds and hybrids, and based on eMPH

The correlation coefficients in expression variation between different genes were calculated. If we assume that the correlation between gene expression and trait performance provide information useful for the identification of the genes controlling the trait according to the central dogma, then the correlation between the variations of different gene expressions would provide information useful for construction of the gene network for the trait of heterosis. According to this hypothesis, our laboratory previously developed a computer program named the GeneNet. The degree of the correlation in variation between different gene expressions and between the gene expression and phenotypes in a population would provide useful information to discover genes controlling phenotypic traits and predict the gene functional network.

3.7.1 Pairwise correlation and network of gene expression in inbred lines

A vast majority of traits, if not all, particularly those of complicated quantitative traits, result from the expression interaction or cooperation of many genes in a manner of either networks or complex formation. Therefore, we inferred the working or expression relationships between the genes by expression correlation analysis using the multivariate test with JMP v8.0. The genes having a correlation coefficient value ≥ 0.6 with P value ≤ 0.05 were considered to be significantly related. In this study, only in the case of the correlation with a coefficient ≥ 0.8000 and a significance level of $P \leq 0.05$, the pair of genes was considered to be highly related in transcription activity. Figure 12 shows the network of 39 genes at a threshold of 0.6, 0.7, 0.8 and 0.9, respectively. (Table 6) When

the correlation efficient was set at 0.6 or larger, 33 of the 39 genes (84.62%) formed a network (Figure 12A). When the correlation efficient was set at 0.7 or larger, 21 of the 39 genes (53.85%) remained in the network (Figure 12B). When the correlation efficient was 0.8 or larger, 10 of the 39 genes (25.64%) were constructed into two independent networks, one containing two genes and the other containing 8 genes (Figure 12C). When the correlation efficient was 0.9 or larger, three independent networks formed, with each containing two genes (Figure 12D). They were 2D3 (chlorophyll a-b binding protein) with 2D4 (chlorophyll a-b binding protein), 3D2 (At1g69640/F24J1.22) with 3U10 (putative branched chain alpha-keto acid dehydrogenase E2 subunit), and 2D6 (histone H4) with 2D9 (unknown protein), indicating that these genes must play important roles in pair in the network or complex of a biological process. Additionally, since 2D4 and 2D3 share the same functions in annotation, the formation of the network between them provided a strong validation of the method in the network construction.

3.7.2 Pairwise correlation and network of gene expression in hybrids

Similarly, we constructed the network of the 39 genes in expression in the F_1 hybrids. Figure 13 shows the gene network at a threshold of 0.6, 0.7, 0.8 and 0.9, respectively. (Table 7) When the correlation coefficient was set at 0.6 or larger, 32 of the 39 genes (82.05%) formed a network (Figure 13A). When the correlation coefficient was set at 0.7 or larger, 24 of the 39 genes (61.54%) were remained in the network (Figure 13B). When the correlation coefficient was at 0.8 or larger, 10 of the 39 genes (25.64%) were constructed into three independent networks, with one containing two genes, the second one containing three genes and the third one containing five genes (Figure 13C).

Although four of the 10 genes were shared with the network of 8 genes at the same stringency (0.8) in inbreds, the network of the five genes newly formed in the hybrids. When the correlation coefficient was set at 0.9 or larger, only 3 of the 39 (7.69%), 1U4 (unknown), 1U5 (unknown) and 1U10 (unknown), formed a network (Figure 13D), with none of them sharing with those constructed at the same stringency in inbreds, suggesting their cooperation and correlation in a biological process. The differences of the gene networks between inbreds and their hybrids strongly suggest the roles of the gene network alternations in heterosis form inbreds to hybrids.

3.7.3 Pairwise correlation and network of genes in eMPH

Figure 14 shows the network of the genes constructed based on eMPH at a threshold of 0.6, 0.7, 0.8 and 0.9, respectively. (Table 8) When the correlation coefficient was set at 0.6 or larger, 37 of the 39 genes (94.87%) formed a network (Figure 14A). When the correlation coefficient was set at 0.7 or larger, 34 of the 39 genes (87.18%) remained in the network (Figure 14B). When the correlation coefficient was at 0.8 or larger, 18 of the 39 genes (46.15%) formed two networks, one containing 6 genes and the other containing 12 genes (Figure 14C). The number and set of genes that formed networks was nearly the combinations of those in inbreds and hybrids. When the correlation coefficient was set at 0.9 or larger (Figure 14D), 7 of the 39 genes formed three networks, with two having 2 genes per network and one having three genes. The networks included 1U5 with 1U4 and 1U9; 3U10 (putative branched chain alpha-keto acid dehydrogenase E2 subunit) with 3D2 (Bax inhibitor-1) and 2D9 (unknown protein) with 2D6 (nucleasome/chromatin assembly factor D). The network of 1U4, 1U5 and

1U9 was present in hybrids while the networks of 3U10 with 3D2 and 2D9 with 2D6 were present in the inbreds. Nevertheless, the network of 2D3 with 2D4 observed in the inbreds was not in the gene networks constructed based on eMPH.

3.8 Pairwise correlation and network between gene expression and phenotypic traits among 21 hybrids

To infer the functions of the genes and establish the gene networks for grain yield trait development, we calculated the correlation between gene expression and trait performance, and constructed the gene network, as we did for the genes, for each trait that was significantly correlated with the expression of one or more of the genes. The expression of the genes was analyzed against the 13 phenotypic traits collected in two biological replicates. Of the 39 genes studied, the expressions of 36 were significantly correlated with the variation of one or more of 12 of the 13 traits (Figure 15) (Table 9). Furthermore, 22 of the 36 genes were mapped to the QTL intervals of maize grain yield traits (Qin et al. 2010), which further verified the results obtained in this study. These traits included days to pollens shedding (DTP), percentage of stalk lodging (STL), percentage of root lodging(RTL), plant height in cm (PHT), ear height in cm (EHT), stay green rating on AUG 22 (SG822), husk coverage rating (HUSK), corn earworm feeding damage (CEW), ear length (EL), plants with common smut (%) (SMUT), percentage of molded kernels (MOLD), plot yield (PLOTYLD), and grain moisture at harvest (MOIST).

Days to pollens shedding (DTP) (Figure 16): When the correlation coefficient threshold was set at a level of 0.3, 10 genes (3U9, 3D2, 1D4, 1D9, 2D10, 1D3, 2D1,

3D1, 2D6, 2D9) and four traits (EHT, MOIST, PHT and HUSK) were correlated and formed a network with DTP. When the threshold was at a level of 0.4, genes 2D6 (nucleasome/chromatin assembly factor D), 3D1 (putative serine/threonine protein kinase), 3U9, 1D4, 1D9, 1D3, and traits PHT and HUSK were still wired with DTP. At the threshold of ≥ 0.5, only gene 3U9 (unknown) and 1D9 (proliferating cell nuclear antigen) were connected with DTP. Previous study showed that the 1D9 gene, as PCNA, actively expresses in tapetum (Hobo et al., 2008) and tapetal cells up to the day before pollen shedding (Rowley et al., 2000), indicating the role of the 1D9 gene in "days to pollen shedding". These previous studies verified our results from gene-trait network construction. Therefore, the genes 3U9 and 1D9 as well as the genes 2D6, 3D1, 1D4 and 1D3 must play important roles in "days to pollens shedding".

Plant height (PHT) (Figure 17): At the correlation coefficient threshold level of 0.3, genes 3D7 (putative diphosphate-fructose-6-phosphate 1-phosphotransferase alpha chain), 2U1 (putative auxin-repressed protein), 1D4, 1D9 (proliferating cell nuclear antigen) and 2D6, and traits EHT and DTP were correlated with PHT and formed a network. At the level above 0.5, genes 1D4 (OSJNBb0072M01.18) and 1D9 (proliferating cell nuclear antigen), and trait EHT were correlated with PHT and remained in the network. Therefore, genes 1D4 and 1D9, and genes 3D7, 2U1, and 2D6 likely significantly contributed to the performance of the trait "Plant height".

Ear height (EHT) (Figure 18): At a correlation coefficient level of 0.3, 1D4, 1D9, 3D7 (putative diphosphate-fructose-6-phosphate 1-phosphotransferase alpha chain), 3U4 (protein kinase homolog), 1U2 (Zinc-finger protein), PHT, DTP, STL, SG822 and EL

were correlated with EHT and was constructed into a network. At a level of 0.4 - 0.5, only trait PHT was in the network with EHT. Therefore, genes 1D4, 1D9, 3D7, 3U4, and 1U2 may be involved in development of trait EHT. The correlation between plant height and ear height, as expected, provided an additional indication of the power of the method used in identification of genes and construction of network for a trait.

Stay green rating on August 22 (SG822) (Figure 19): At a correlation coefficient level of 0.3, 1U7 (unknown), STL, EL, E10W, DTP, MOIST were positively correlated with SG822 whereas the genes 1D10 (contains EST AU164600), 3D6 (histone H4), 3D2 (At1g69640/F24J1.22), 3D3 and 3D5 (putative enoyl-ACP reductase) were negatively correlated with SG822. These genes and traits formed a network with SG822. At the level of correlation coefficient ≥ 0.4, 1U7 and MOIST were positively correlated with SG822 and 3D3 was negatively correlated with SG822, only two genes and one trait staying in the network. When the correlation coefficient was set at ≥ 0.5, only 1U7 was wired with SG822. Therefore, 3D3 (minichromosomal maintenance factor) and 1U7 significantly contributed to the trait stay green while 3D3 acted as a negative factor of the trait development.

Husk coverage (HUSK) (Figure 20): At a correlation coefficient level of 0.3, 1D4, 1D3, 3U9, 1U5, 3D3, 1U9, 2U2, 1U4, 2U3, DTP, and CEW were correlated with HUSK. At the level of correlation coefficient = 0.4 to 0.5, 1D4, 1D3 (unknown) and DTP are related with HUSK. Therefore, 1D3 and 1D4 significantly contributed to the husk development, and the days to pollen shedding was related with the husk coverage in development.

Corn earworm feeding damage (CEW) (Figure 21): At a correlation coefficient level of 0.3, 1D6 (beta-expansin), EL, PlotYLD, SMUT, HUSK and MOIST were correlated with CEW. 3U5 (putative protein), 1U5 (NA), 2U2 (putative NAC-domain protein), 2U3 (drought-induced hydrophobic protein), 1U9 (NA) and 1U4 (NA) were negatively correlated with CEW. At the level of correlation co-efficient = 0.4 to 0.5, only gene 1D6 (beta-expansin) was related with CEW. This result indicated that gene 1D6 was highly correlated with CEW, while 3U5, 1U5, 2U2, 2U3, 1U9 and 1U4 may offer resistance to corn earworm feeding damage.

Ear length (EL) (Figure 22): At a correlation coefficient level of 0.3, 3U10 (putative branched chain alpha-keto acid dehydrogenase E2 subunit), E10W, EHT (ear height in cm), SG822 and CEW formed a network with EL. At the level of correlation coefficient ≥ 0.4, 3U10 and E10W were still in the network with EL, suggesting that 3U10 played an important role in the development of trait "ear length". At the level of correlation coefficient ≥ 0.5, only E10W was connected to EL, indicating their close relationship in development.

Plants with common smut (%) (SMUT) (Figure 23)*:* At a correlation coefficient level of 0.3 to 0.4, 3U9 1U7, CEW and MOIST were networked with SMUT while1U7 and CEW were negatively correlated with SMUT. Therefore, 3U9 and 1U7 may be involved in the development of SMUT, particularly 1U7 may offer resistance to corm smut.

Percentage of molded kernels (MOLD) (Figure 24): At a correlation coefficient level of 0.3, 1D3, 1U2 (Zinc-finger protein 1), 3U2 (Bax inhibitor-1) and 3D7 (putative

diphosphate-fructose-6-phosphate 1-phosphotransferase alpha chain) formed a network with MOLD. At a correlation coefficient level of 0.4 to 0.5, 1U2 and 3U2 remained in the network with MOLD while 3U2 was negatively correlated with MOLD. This result suggests that both 1U2 and 3U2 played significant roles in kernel mold; 3U2 may offer resistance to corn kernel mold infection.

Plot yield (PLOTYLD) (Figure 25): At a correlation coefficient level of 0.3, 16 genes and CEW formed a network with PLOTYLD, while 10 of them were negatively correlated with PLOTYLD. At the level of correlation coefficient = 0.4 to 0.5, nine of the genes were maintained in the network with PLOTYLD, of which five were negatively correlated with PLOTYLD. At the level of correlation coefficient above 0.5, 2D4 (chlorophyll a/b binding protein), 2U2 (putative NAC-domain protein) and 2U3 (drought-induced hydrophobic protein) remained in the network with PLOTYLD, with only 1U9(NA) being negatively correlated with PLOTYLD. This indicates that grain yield is a consequence of balanced interaction between a number of genes, including positive and negative, in which2D4, 2U2, 2U3 and 1U9 might play more important roles than others.

Grain moisture at harvest (MOIST) (Figure 26): At a correlation coefficient level of 0.3 to 0.4, 10 genes, including 4 positive- and 6-negatively related genes, and four traits (SMUT, SG822, DTP and CEW) formed a network with MOIST. At the level of correlation coefficient = 0.4 to 0.5, 1U7, 3D5 and SG822 stayed in the network with MOIST. At the level of correlation coefficient above 0.5, 3D5 (putative enoyl-ACP reductase) and 1U7were still wired with MOIST, though 1U7 was negatively correlated

with MOIST. This result indicates that both 3D5 and 1U7 played an important role in corn grain moisture at harvest, but 3D5 played a positive role while 1U7 acted as a negative regulator.

Stalk lodging (STL) (Figure 27): At a correlation coefficient level of 0.3, 1D6 (beta-expansin), 2D4 (chlorophyll a/b binding protein) and SG822 formed a network with STL. This indicates that stalk lodging are related with stay green and regulated by the activities of 1D6 and 2D4.

CHAPTER IV

DISCUSSION AND CONCLUSIONS

Genome research is entering into the post-genome era. One of the most challenging tasks in the era is to identify the genes controlling trait development and determine the molecular basis of the trait performance, especially those of complicated, quantitative traits because no effective strategies have been developed for their advanced studies. In this study, we have demonstrated a new, rapid and powerful strategy to identify genes contributing to or responsible for and construct gene network for development and performance of a complicated quantitative trait using maize heterosis in grain yield traits and a panel of selected candidate genes as the model system. By including the genes of known function and network, and genes unrelated as the controls of the study, the strategy has been well validated in this study. Using the strategy, we have constructed a number of genes contributing to different traits of importance to grain yield and constructed their initial network (see below, for more discussion). Therefore, the strategy will be applicable for identifying the genes and constructing networks of traits in other species at a genome-wide scale. Furthermore and importantly, this study has led to several significant findings that are crucial to our better understanding of maize heterosis in the grain yield traits.

4.1 The expression networks of genes have been alternated from inbreds to their F_1 hybrids

We have, for the first time, constructed the networks from some of the selected genes that most differentially expressed in developing ear shoots between a super F_1 hybrid and its parental inbreds, Mo17 and B73 (Qin et al., 2010), in both a set of inbreds and their F_1 hybrids. The network formation from some of the selected genes indicates that the genes that actively express simultaneously in a same organ may act correlatively for a trait development or are involved in a same or related biological process. Comparative analysis has revealed that the network of the genes have been alternated significantly from inbreds to their F_1 hybrids For instance, at the threshold of correlation coefficient ≥ 0.8, a high stringent cutoff, 10 of the 39 genes studied formed one network in inbreds whereas 10 of the 39 genes studied formed three independent networks in hybrids. Of the 10 genes in the networks, only four were shared between inbreds and hybrids. At the threshold of correlation coefficient ≥ 0.9, six of the 39 genes studied formed three independent networks in inbreds whereas only one network formed from three genes in hybrids. The networks in the inbreds disappeared in the hybrids; instead of those, a new network formed in the hybrids. Furthermore, when the eMPH was used as a parameter for the gene network construction at the threshold of correlation coefficient $\geq$ 0.8, 18 of the 39 genes formed two networks. One network contained 12 of the 18 genes, of which nine were from the 10 genes in the two networks of inbreds and five from the five genes constituting two of the three networks of hybrids. The network of the remaining 6 genes constructed from eMPH contained all six genes of the third network

in hybrids, which was not present in inbreds. These results suggest that the altered gene networks in the hybrids relative to their parental inbreds may play an important role in heterosis. Crossing diverged inbreds that brings two sets of genes into a new genetic system, including gene or allele interactions, leads to significant changes in gene expression networks, which in turn forms the molecular basis of trait heterosis. Such gene networks changes from inbreds to hybrids could result from gene actions in a manner of additivity, dominance, overdominance, pseudo-overdominance, epistasis, and/or their combination since a number of genes are likely involved in a trait heterosis. Heterosis of a trait is the consequence of balancing of these gene actions.

4.2 The genes have different expression patterns between inbreds and their hybrids, but they vary from hybrid to hybrids

It has been observed in previous studies that many genes are altered in expression patterns when two inbreds are crossed to yield F_1 hybrids (Bao at al 2005; Swanson-Wagner et al. 2006; Wei et al. 2009; Thiemann et al. 2010; Qin et al. 2010). As expected, the cluster analysis of the inbreds, hybrids, and hybrids with the mean of the two parents supports the previous conclusion (Figures 7 - 10). It seems consistent among different studies (Bao at al 2005; Swanson-Wagner et al. 2006; Wei et al. 2009; Thiemann et al. 2010; Qin et al. 2010) that differential expressions of genes between inbreds and hybrids are the molecular basis of trait heterosis. However, this study reveals that the differential expressions, as presented in eMPH, vary significantly in the same organ (developing ear shoots) at the same developmental stage (V13) among different F_1 hybrids. That the eMPHs of some of the 39 genes studies were more conservative than

those of others across different hybrids may indicate the importance of the genes in the general basis of heterosis, which may be somehow related to traditional general combination ability. On the other hand, the genes having more variable eMPH across hybrids may be of indication for traditional special combination ability. Further studies will be needed to determine what kind of gene expression petterns will yield super heterotic hybrids that could be used in agricultural production.

4.3 A trait performance is resulted from the interaction of multiple genes that form a network

Genetic correlation analysis and gene-trait network construction have allowed identification of some genes contributing to different trait performances from the 39 genes studied. We have examined the 13 traits that are significantly related to maize grain yield since our mRNA tissues used in this study were the developing whole ear shoots that are considered to be significant for grain yield. Despite of the fact that only 39 genes most differentially expressed between F_1 hybrid (M017 x B73) and their parental inbreds were selected for and used in the study, we have identified 1 – 16 genes per each of 12 of the 13 traits from the gene pool that were significantly correlated and formed networks with the traits. (Table 10)This indicates that multiple genes are involved in a trait development, or the performance of a trait results from interaction of multiple genes. How many genes are involved in a trait development may depend on the complexity of the trait. For instance, we have identified 16 genes for grain yield, 10 genes for grain moisture and 10 genes for days to pollen shedding, but only one gene for ear length. Finally, it should pointed out that because only selected 39 genes were used

in this study, the gene networks for the trait performances constructed in this study are initial or partial. A genome-wide research will be needed to complete the gene network for each trait. At the same time, additional genes will be expected to be identified for each trait; hence, a more complicated gene network for a trait performance will be expected.

In total, we have identified 36 out of the 39 genes whose expressions are correlated with the variations of the 12 traits. Of the 36 genes that are wired with one or more of the traits, 21 were found in the QTL intervals of grain yield traits previously mapped, such as 1000-kernal weight, cob diameter, kernel row number per ear, kernel row length, days to pollen shedding, plant height, ear height, ear length, vegetative to generative transition, test weight, and grain yield (Qin et al. 2010) (Table 9). Moreover, 8 of the 39 selected genes happened in the list of the genes that were correlated with maize hybrid performance for grain yield, hybrid performance for grain dry matter content and grain yield MPH identified from 9-day seedlings of a panel of maize inbreds (Thiemann et al. 2010). In this study, 7 of the eight genes were found to correlate with and form networks with 10 of the 13 traits studied. These include smut resistance, stay green, grain yield, grain moisture, plant height, ear height, husk coverage, days to pollen shedding, kernel mold resistance and ear length. These results have further validated those of gene identification and network construction in this study.

4.4 Gene expression varies significantly among inbreds, among hybrids and in heterosis

Since gene expression, or mRNA level, is the immediate products of genes, as does a phenotypic trait - the final product of genes, it is expected that variation widely exists in gene expression among different genotypes. This has constituted the molecular basis of phenotypic variation. In this study, we have comprehensively tested and confirmed this hypothesis using 13 inbred lines and 20 hybrids of maize. Our study shows all of the 39 genes studied display significant variations in expression (P-value ≤ 0.05) in the developing ear shoots among either the inbreds or the hybrids. The variations have been further verified by the hierarchical clustering heat maps constructed. This result is consistent with the conclusion made in previous studies that transcriptional difference exists between maize inbred lines (Swanson-Wagner et al. 2006; Thiemann et al. 2010). (Table 11)The transcriptional differences among inbred lines may provide useful information or criterion for parent selection for heterosis breeding (Thiemann et al. 2010).

In the eMPH variation analysis, 29 of the 39 genes (74.4%) showed significant variations in expression among the 20 combinations studied, while the remaining 10 genes showed no significant variation. The significant variation of eMPH for a gene is an indication of its inconsistent expression across different combinations, which may include up-regulated in some combinations, while down-regulated in the others. The fact that the 10 genes (3U9, 1U5, 1U7, 1U9, 2D1, 2U3, 2U4, 2U9, 3D2 and 3U10) did not show significant variations may be because those genes are constantly expressed in

different genotypes, including hybrids and inbreds.

REFERENCES

Bao, J.Y., Lee, S.G., Chen, C., Zhang, X.Q., Zhang, Y., Liu, S.Q., Clark, T., Wang, J., Cao, M.L., Yang, H.M., Wang, S.M., and Yu, J. (2005). Serial analysis of gene expression study of a hybrid rice strain (LYP9)and its parental cultivars. Plant Physiol **138,** 1216-1231.

Crow, J.F. (1952). Dominance and overdominance. p. 282-297. *In*: Heterosis J. W. Gowen (ed.). Iowa State College Press, Ames.

Crow, J.F. (1998). 90 years ago: The beginning of hybrid maize. Genetics **148,** 923-928.

Davenport, C.B. (1908). Degeneration, Albinism and Inbreeding. Science **28,** 454-455.

Duvick, D.N. (1999). Heterosis: feeding people and protecting natural resources. Genetics and Exploitation of Heterosis in Crops **524,** 19-29

Duvick, D.N. (2001). Biotechnology in the 1930s: the development of hybrid maize. Nat Rev Genet **2,** 69-74.

Duvick, D.N. (2005). The contribution of breeding to yield advances in maize (Zea mays L.). Adv Agron **86,** 83-145.

Frascaroli, E., Cane, M.A., Landi, P., Pea, G., Gianfranceschi, L., Villa, M., Morgante, M., and Pe, M.E. (2007). Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. Genetics **176,** 625-644.

Goodnight, C.J. (1999). Epistasis and heterosis. Genetics and Exploitation of Heterosis in Crops **524,** 59-68

Guo, M., Rupe, M.A., Yang, X.F., Crasta, O., Zinselmeier, C., Smith, O.S., and Bowen, B. (2006). Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. Theor Appl Genet **113,** 831-845.

Guo, M., Yang, S., Rupe, M., Hu, B., Bickel, D.R., Arthur, L., and Smith, O. (2008). Genome-wide allele-specific expression analysis using Massively Parallel Signature Sequencing (MPSS (TM)) Reveals cis- and trans-effects on gene expression in maize hybrid meristem tissue. Plant Mol Biol **66,** 551-563.

Hobo, T., Suwabe, K., Aya, K., Suzuki, G., Yano, K., Ishimizu, T., Fujita, M., Kikuchi, S., Hamada, K., Miyano, M., Fujioka, T., Kaneko, F., Kazama, T., Mizuta, Y., Takahashi, H., Shiono, K., Nakazono, M., Tsutsumi, N., Nagamura, Y., Kurata, N., Watanabe, M., and Matsuoka, M. (2008). Various spatiotemporal expression profiles of anther-expressed genes in rice. Plant Cell Physiol **49,** 1417-1428.

Hua, J.P., Xing, Y.Z., Wu, W.R., Xu, C.G., Sun, X.L., Yu, S.B., and Zhang, Q.F. (2003). Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. P Natl Acad Sci USA **100,** 2574-2579.

Huang, Y., Zhang, L.D., Zhang, J.W., Yuan, D.J., Xu, C.G., Li, X.H., Zhou, D.X., Wang, S.P., and Zhang, Q.F. (2006). Heterosis and polymorphisms of gene expression in an elite rice hybrid as revealed by a microarray analysis of 9198 unique ESTs. Plant Mol Biol **62,** 579-591.

Li, Z.K., Luo, L.J., Mei, H.W., Wang, D.L., Shu, Q.Y., Tabien, R., Zhong, D.B., Ying, C.S., Stansel, J.W., Khush, G.S., and Paterson, A.H. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. Genetics **158,** 1737-1753.

Lu, H., Romero-Severson, J., and Bernardo, R. (2003). Genetic basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. Theor Appl Genet **107,** 494-502.

Luo, L.J., Li, Z.K., Mei, H.W., Shu, Q.Y., Tabien, R., Zhong, D.B., Ying, C.S., Stansel, J.W., Khush, G.S., and Paterson, A.H. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. Genetics **158,** 1755-1771.

Meyer, S., Pospisil, H., and Scholten, S. (2007). Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. Plant Mol Biol **63,** 381-391.

Rowley, J.R., Skvarla, J.J., and Walles, B. (2000). Microsporogenesis in Pinus sylvestris L. VIII. Tapetal and late pollen grain development. Plant Syst Evol **225,** 201-224.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C-T method. Nat Protoc **3,** 1101-1108.

Stuber, C.W., Lincoln, S.E., Wolff, D.W., Helentjaris, T., and Lander, E.S. (1992). Identification of Genetic-Factors Contributing to Heterosis in a Hybrid from 2 Elite Maize Inbred Lines Using Molecular Markers. Genetics **132,** 823-839.

Stupar, R.M., and Springer, N.M. (2006). Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F-1 hybrid. Genetics **173,** 2199-2210.

Stupar, R.M., Gardiner, J.M., Oldre, A.G., Haun, W.J., Chandler, V.L., and Springer, N.M. (2008). Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. Bmc Plant Biol **8,** -.

Swanson-Wagner, R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D., and Schnable, P.S. (2006). All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci U S A **103,** 6805-6810.

Tang, J.H., Yan, J.B., Ma, X.Q., Teng, W.T., Wu, W.R., Dai, J.R., Dhillon, B.S., Melchinger, A.E., and Li, J.S. (2010). Dissection of the genetic basis of heterosis in an elite maize hybrid by QTL mapping in an immortalized F-2 population. Theor Appl Genet **120,** 333-340.

Valasek, M.A., and Repa, J.J. (2005). The power of real-time PCR. Adv Physiol Educ **29,** 151-159.

Wei, G., Tao, Y., Liu, G., Chen, C., Luo, R., Xia, H., Gan, Q., Zeng, H., Lu, Z., Han, Y., Li, X., Song, G., Zhai, H., Peng, Y., Li, D., Xu, H., Wei, X., Cao, M., Deng, H., Xin, Y., Fu, X., Yuan, L., Yu, J., Zhu, Z., and Zhu, L. (2009). A transcriptomic analysis of superhybrid rice LYP9 and its parents. Proc Natl Acad Sci U S A **106,** 7695-7701.

Xiao, J.H., Li, J.M., Yuan, L.P., and Tanksley, S.D. (1995). Dominance Is the Major Genetic-Basis of Heterosis in Rice as Revealed by Qtl Analysis Using Molecular Markers. Genetics **140,** 745-754.

APPENDIX

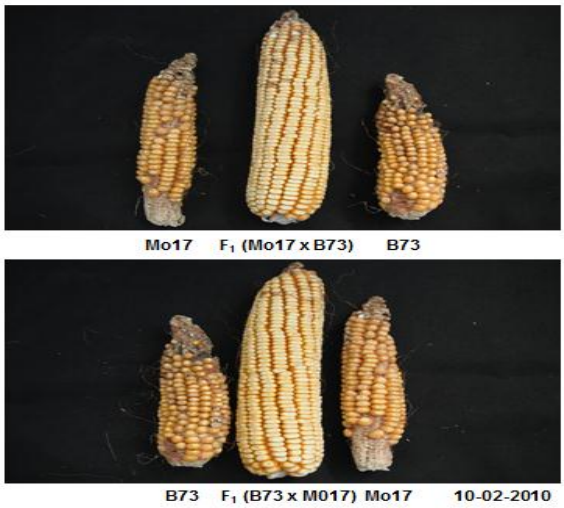


Fig. 1 Heterosis of maize hybrids in phenotypes. F1 hybrid lines (Mo17×B73) and (B73×Mo17) manifest obvious superior performance over either inbred parents Mo17 and B73.

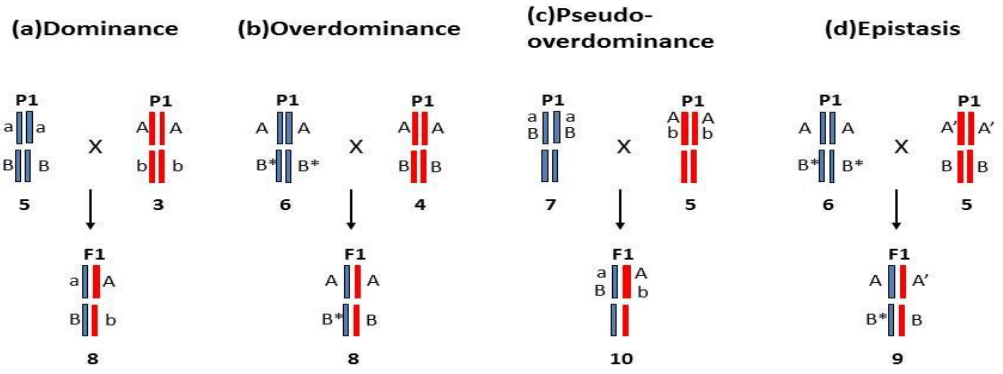


Fig. 2 Genetic models of heterosis. Dominance theory demonstrates that Hybrid vigor is attributed to the complementation of superior dominant alleles from both parental inbred lines at multiple loci in F_1 hybrid over their corresponding unfavorable alleles. Overdominance: Hybrid vigor is attributed to the allelic interaction of alleles from the two parental inbred lines at a locus in F_1 hybrid. Pseudo-overdominance: Hybrid vigor is attributed to the repulsion or *trans* linkage of recessive alleles with dominant alleles from the two parental inbred lines at multiple loci in F_1 hybrid. Epistasis: Hybrid vigor is attributed to the interaction of non-allelic genes from the two parental inbred lines at two or multiple loci in F_1 hybrid.

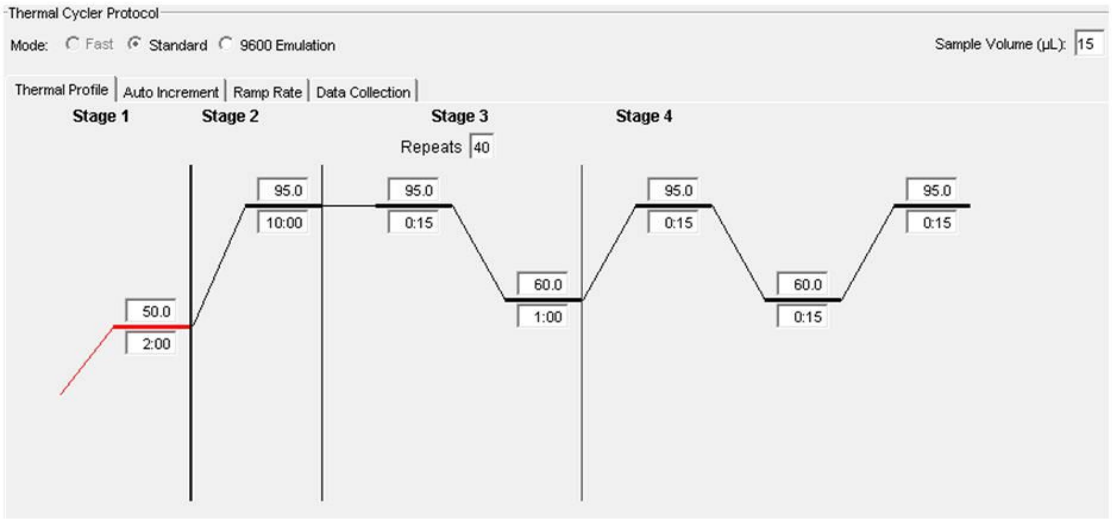


Fig. 3 Reaction condition for RT-qPCR by SDS version 2.3(ABI)

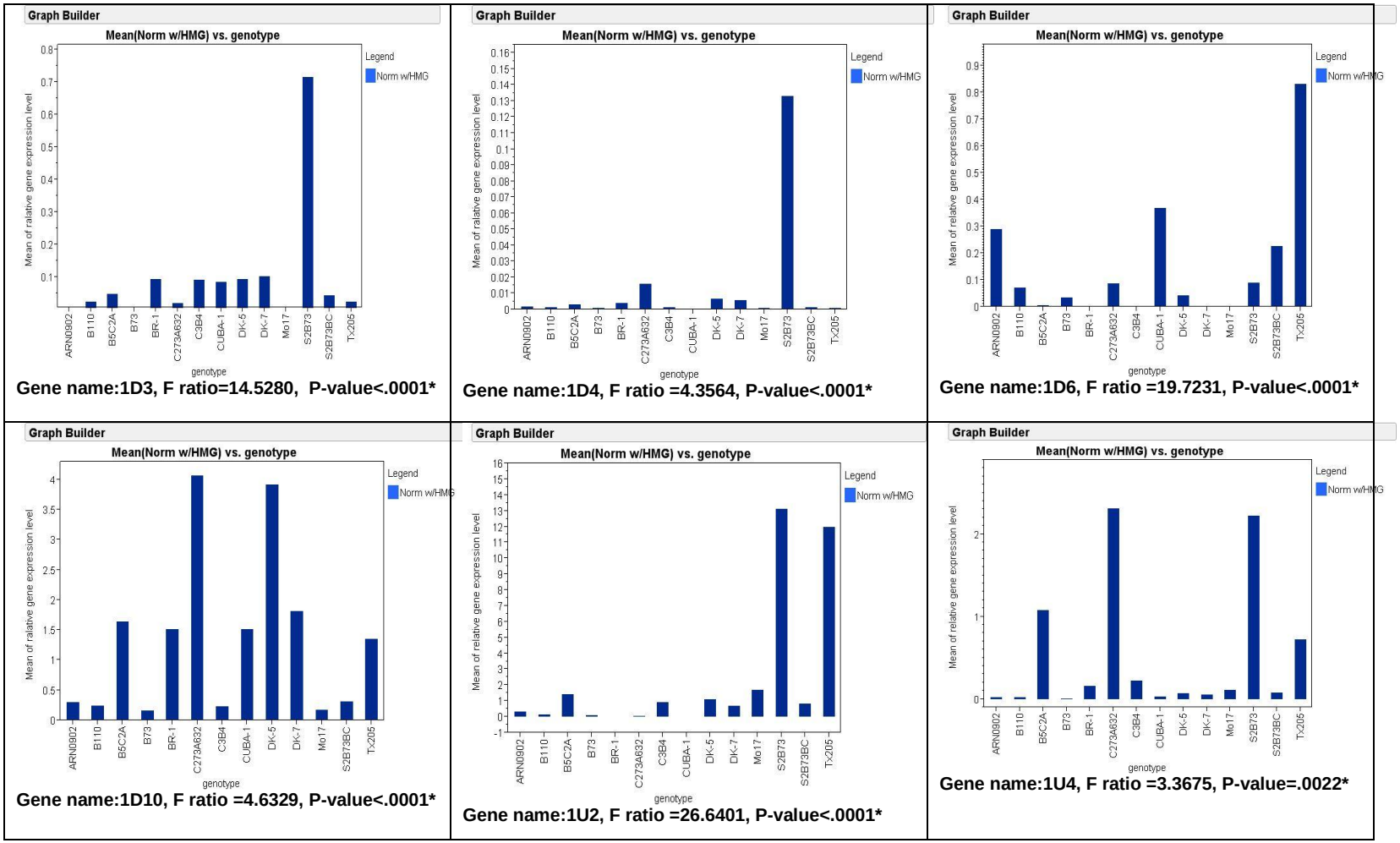
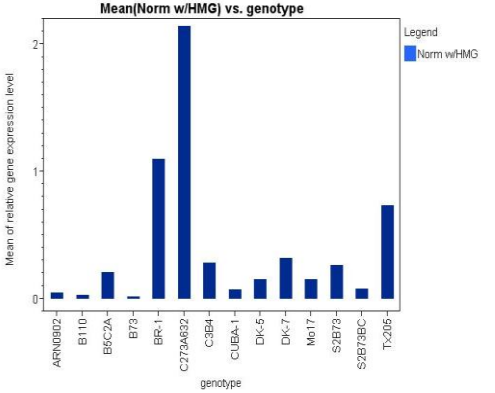
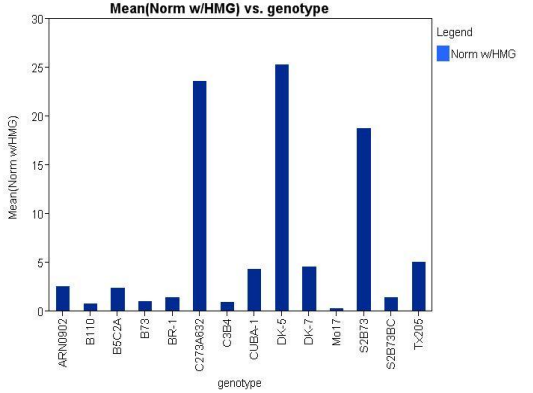


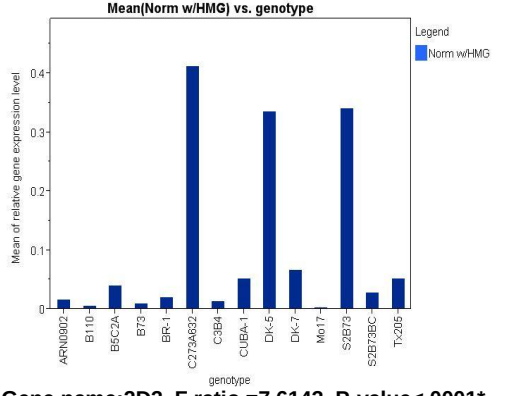
Fig. 4 Gene expression variation of 39 genes among 13 genotypes of inbred lines. The X-axis of each figure represents different genotypes and Y-axis represents the mean of gene expression level derived from the formula $2^{-\Delta T}$.This figure shows that expression of 39 genes varied significantly among 13 inbreds(P-value≤0.05)



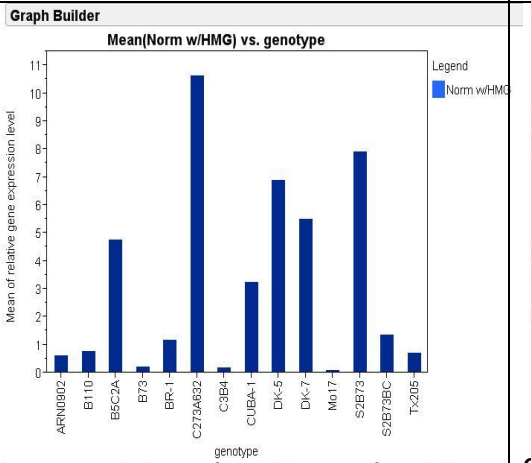
Gene name:1U9,F ratio =3.8398, P-value=0.0005*



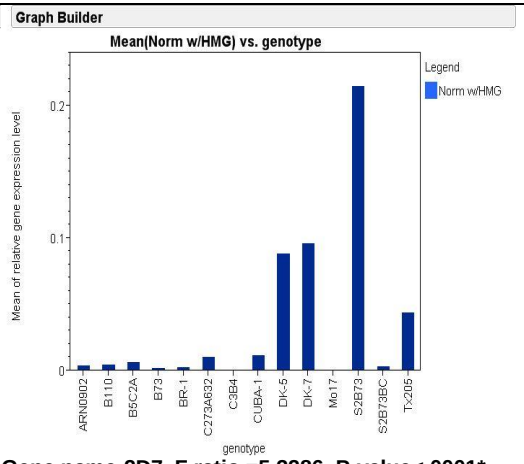
Gene name:2D1, F ratio =4.4468, P-value<.0001*



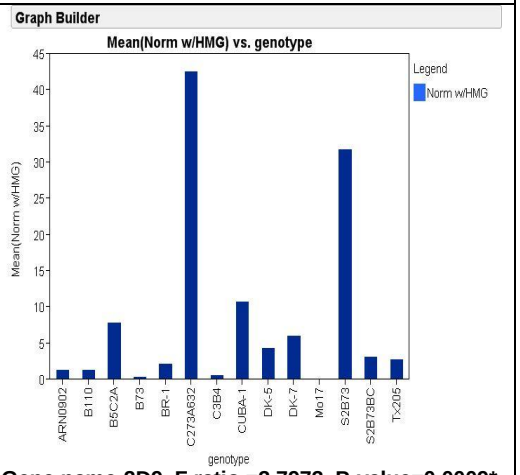
Gene name:2D3, F ratio =7.6143, P-value<.0001*



Gene name:2D6, F ratio =6.0875, P-value<.0001*

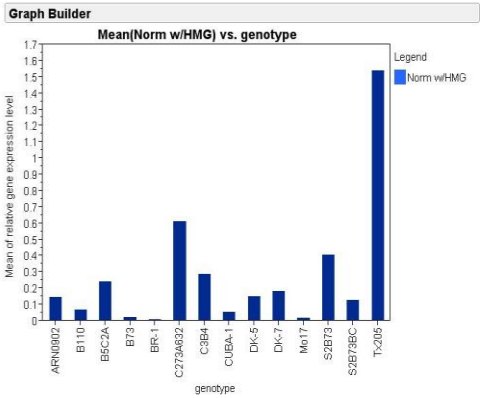


Gene name:2D7, F ratio =5.3286, P-value<.0001*

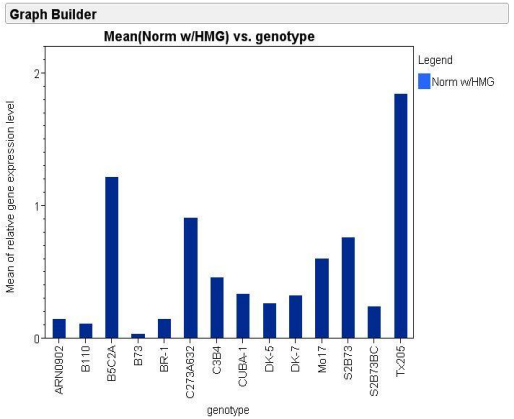


Gene name:2D9, F ratio =3.7973, P-value=0.0009*

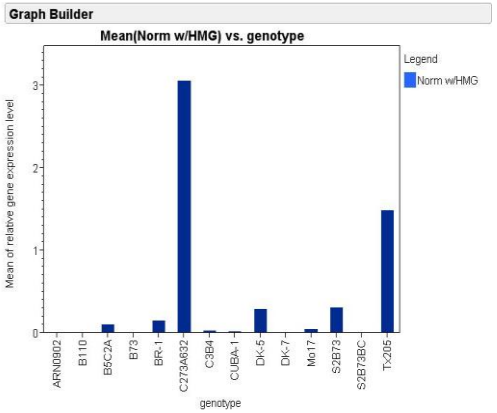
Fig. 4 (Continued)



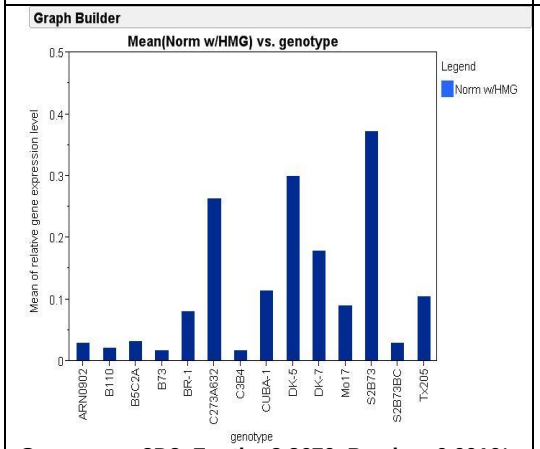
Gene name:2U2, F ratio =7.9070, P-value<.0001*



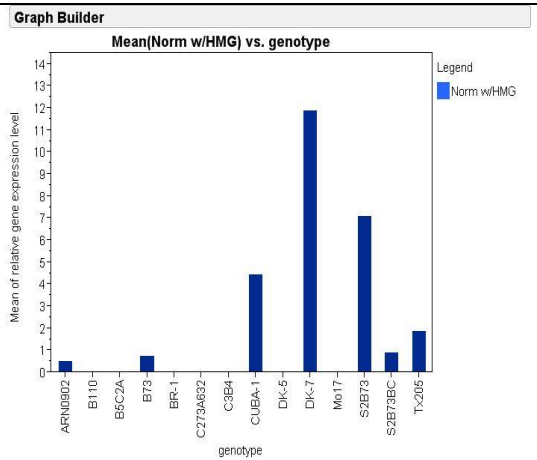
Gene name:2U3, F ratio =2.2588, P-value=0.0263*



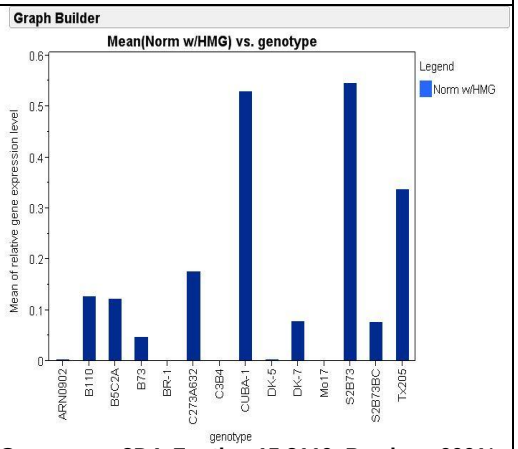
Gene name:2U4, F ratio =25.3106, P-value<.0001*



Gene name:3D2, F ratio=3.3670, P-value=0.0016*

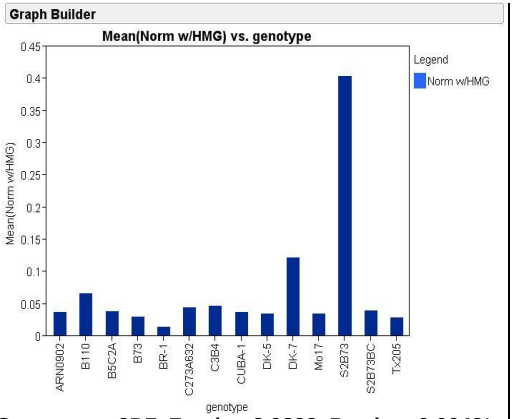


Gene name:3D3, F ratio =28.3329, P-value<.0001*

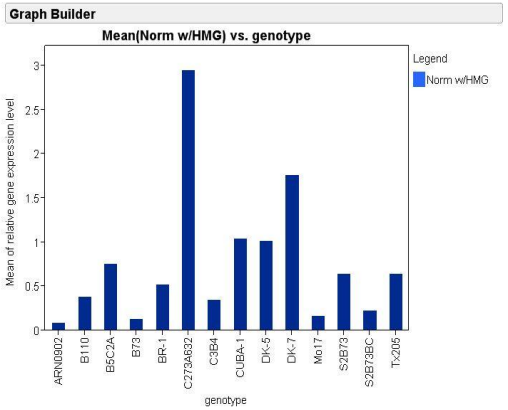


Gene name:3D4, F ratio =15.3110, P-value<.0001*

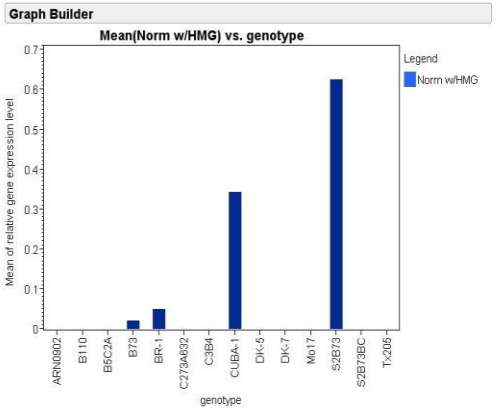
Fig. 4 (Continued)



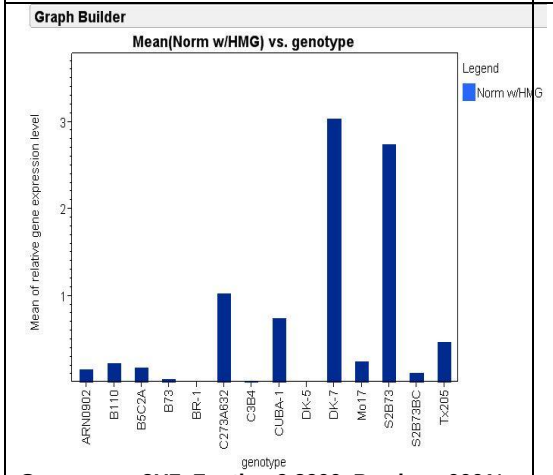
Gene name:3D7, F ratio =2.9822, P-value=0.0042*



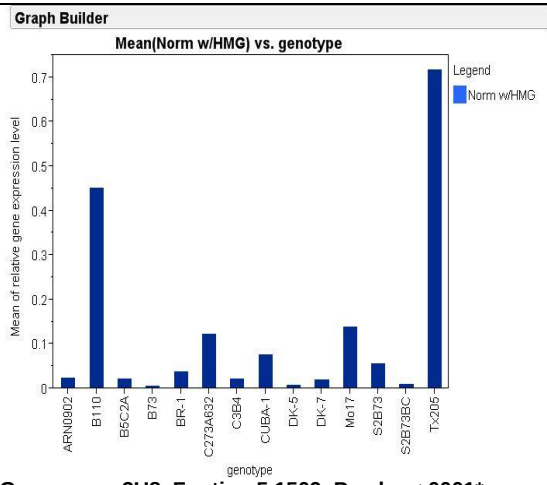
Gene name:3D9, F ratio =3.0793, P-value=0.0033*



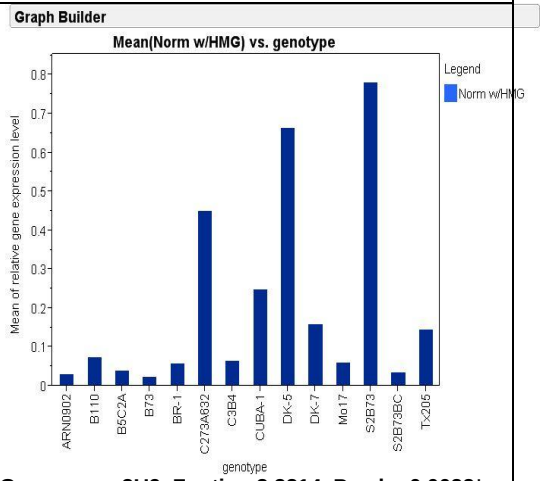
Gene name:3U2, F ratio =21.9086, P-value<.0001*



Gene name:3U7, F ratio =9.2900, P-value<.0001*

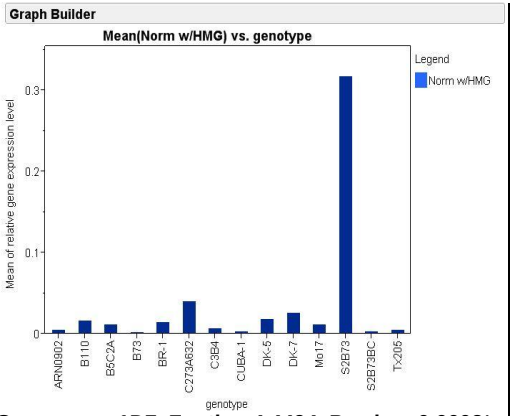


Gene name:3U8, F ratio =5.1509, P-value<.0001*

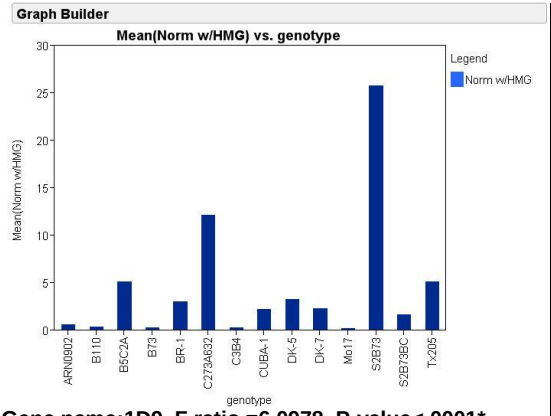


Gene name:3U9, F ratio =3.2314, P-value0.0022*

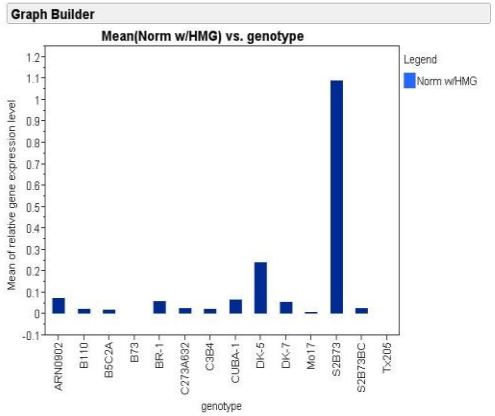
Fig. 4 (Continued)



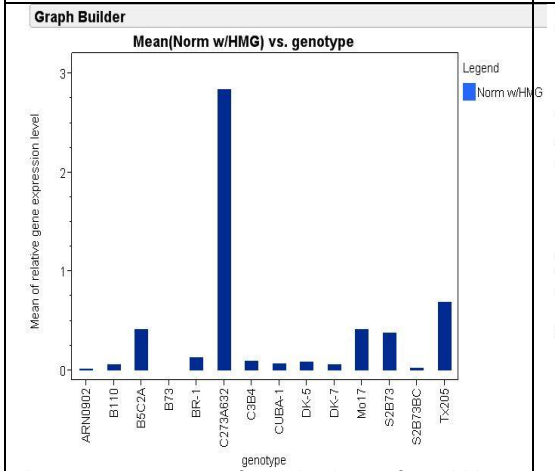
Gene name:1D7, F ratio =4.4494, P-value=0.0002*



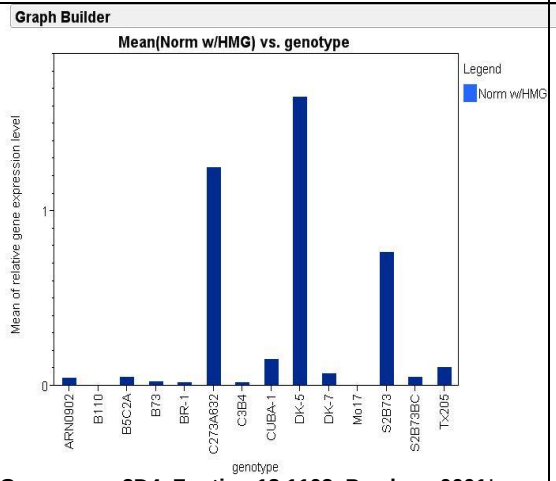
Gene name:1D9, F ratio =6.0978, P-value<.0001*



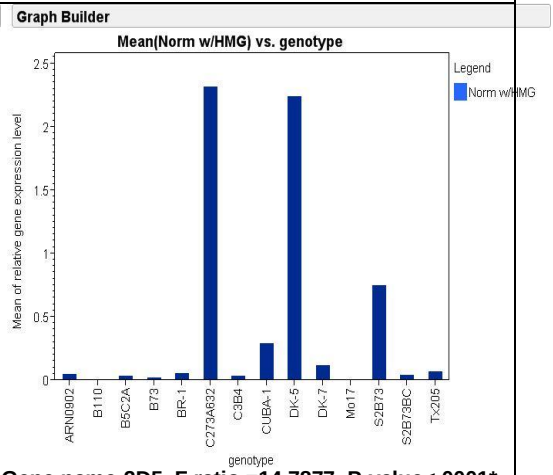
Gene name:1U7, F ratio =6.0530, P-value<.0001*



Gene name:1U5, F ratio =6.3773, P-value<.0001*

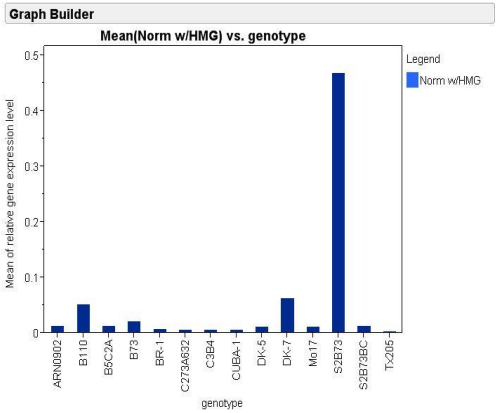


Gene name:2D4, F ratio =18.1102, P-value<.0001*

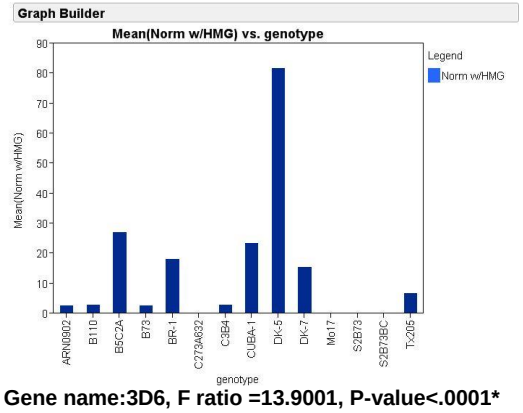


Gene name:2D5, F ratio =14.7877, P-value<.0001*

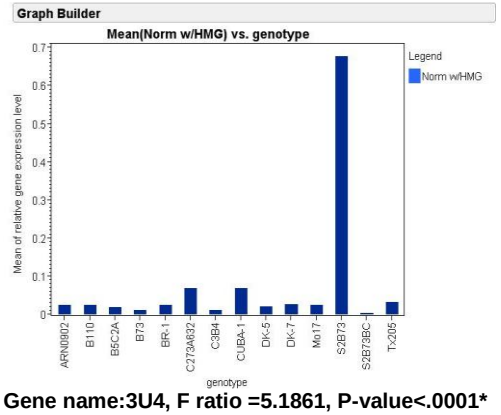
Fig. 4 (Continued)



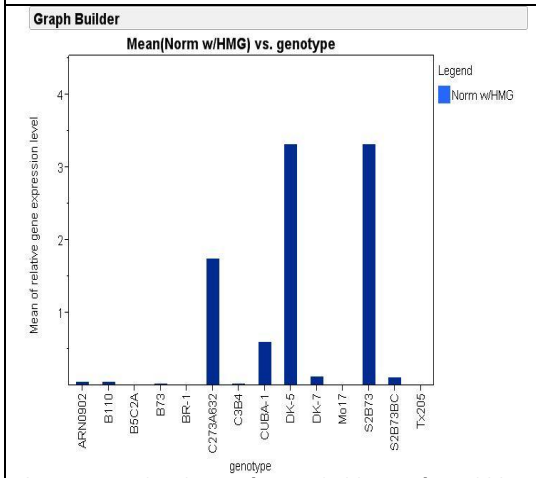
Gene name:3D5, F ratio =6.8680, P-value<.0001*



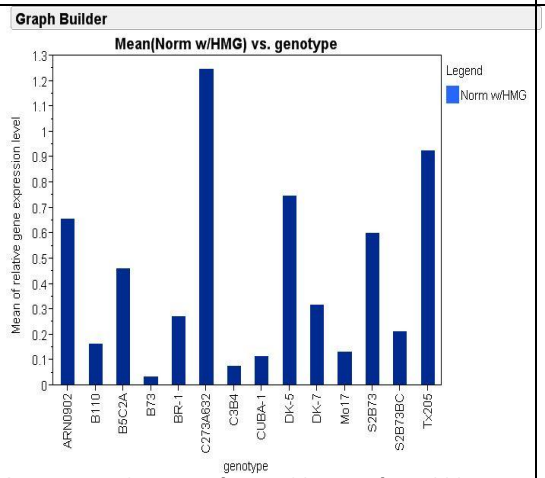
Gene name:3D6, F ratio =13.9001, P-value<.0001*



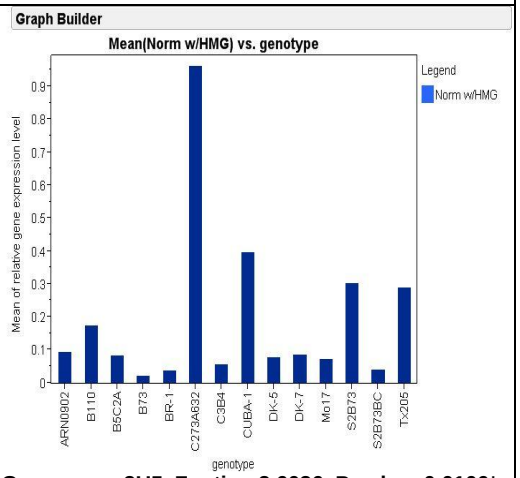
Gene name:3U4, F ratio =5.1861, P-value<.0001*



Gene name:2D10, F ratio =14.6700, P-value<.0001*

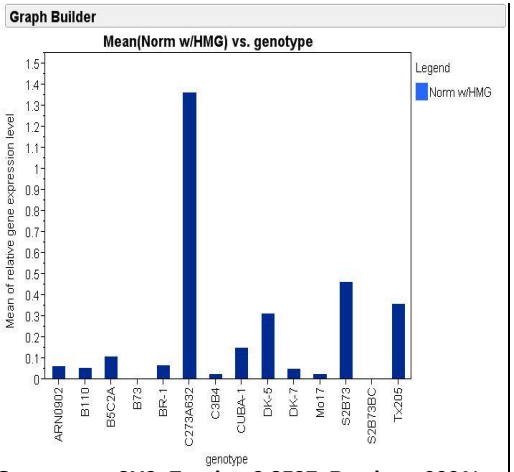


Gene name:2U1, F ratio =5.1305, P-value<.0001*

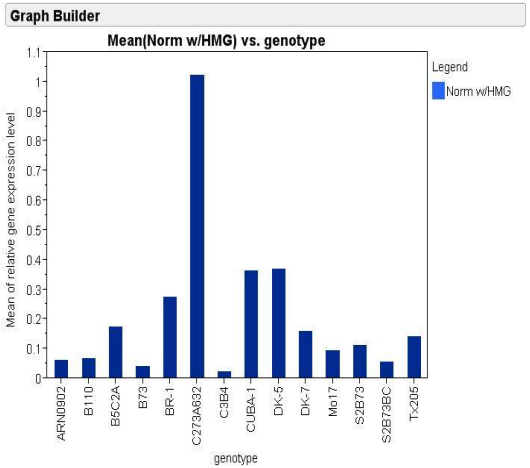


Gene name:3U5, F ratio =2.6036, P-value=0.0106*

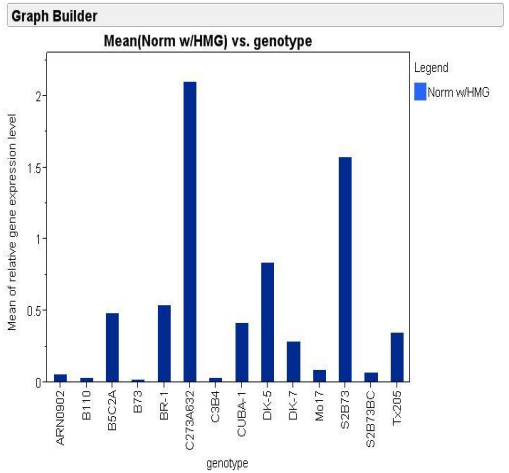
Fig. 4 (Continued)



Gene name:2U9, F ratio =9.2597, P-value<.0001*



Gene name:3D1, F ratio =3.0935, P-value=0.0034*



Gene name:3U10, F ratio =2.3617, P-value<0.0187*

Fig. 4 (Continued)

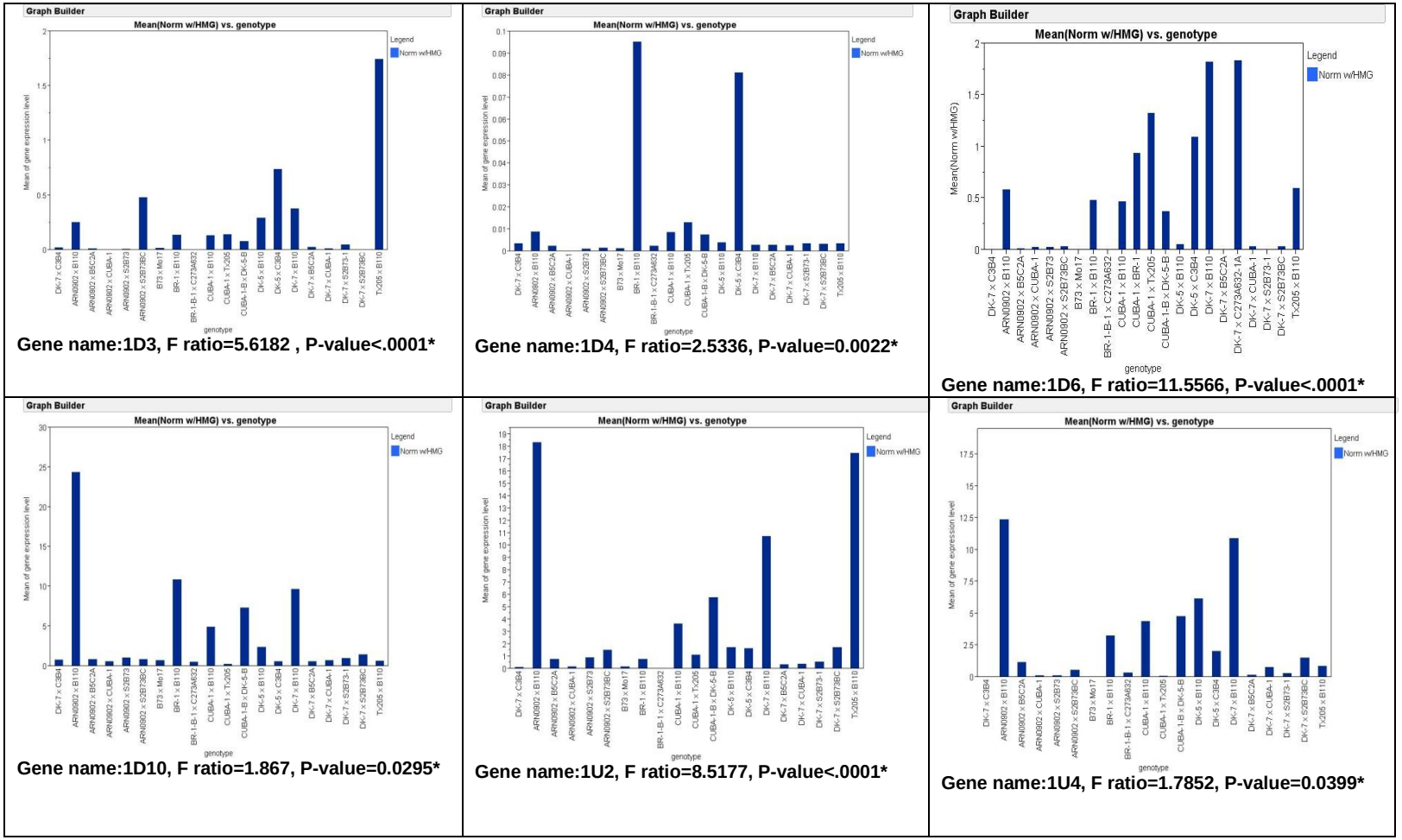
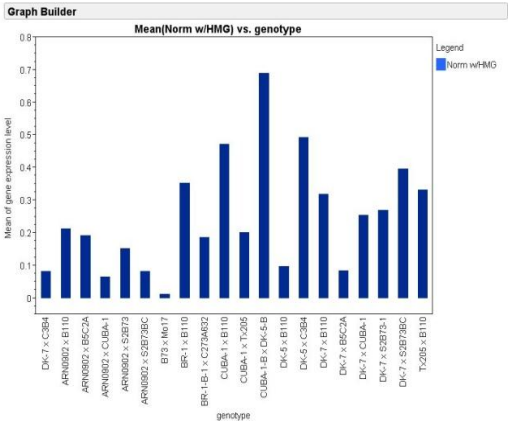
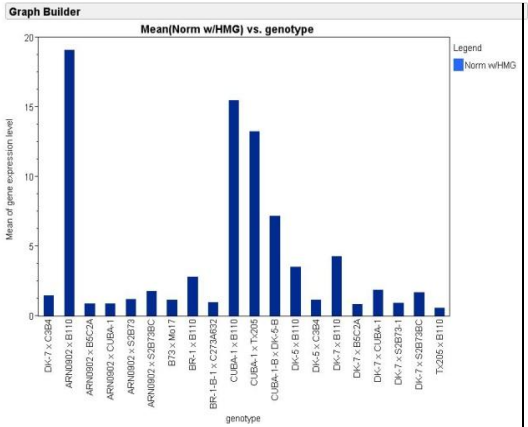


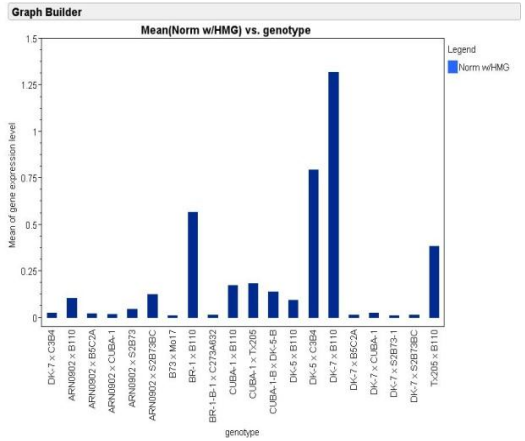
Fig. 5 Gene expression variation of 39 genes among 20 F1 hybrids. The X-axis represents different genotypes of hybrids and Y-axis represents the mean of gene expression level derived from the formula $2^{-\Delta T}$. This figure shows that expression of 39 genes varied significantly among 20 hybrids (P-value≤0.05)



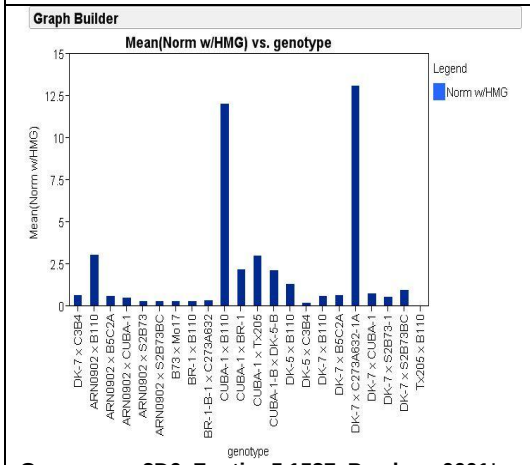
Gene name:1U9, F ratio=4.5609, P-value<.0001*



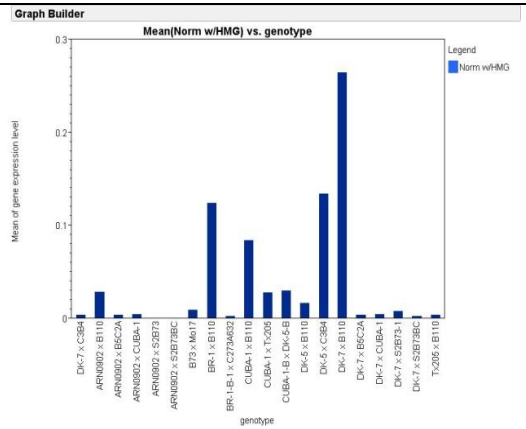
Gene name:1D1, F ratio=2.0625, P-value=0.0142*



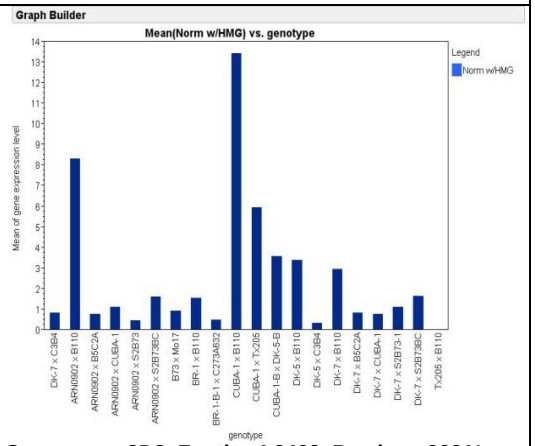
Gene name:2D3, F ratio=3.6261, P-value<.0001*



Gene name:2D6, F ratio=5.1587, P-value<.0001*

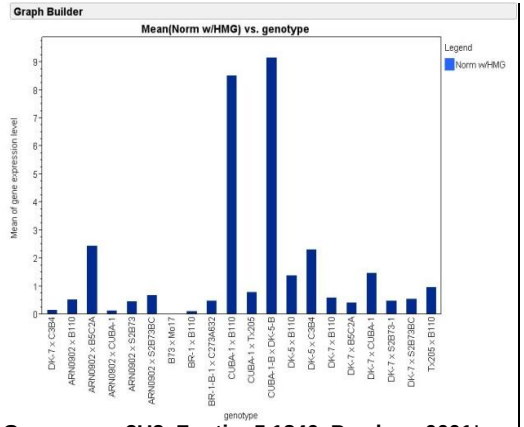


Gene name:2D7, F ratio=2.5508, P-value=0.0020*

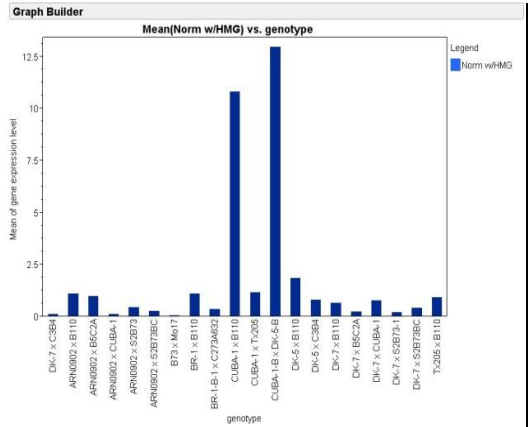


Gene name:2D9, F ratio=4.6460, P-value<.0001*

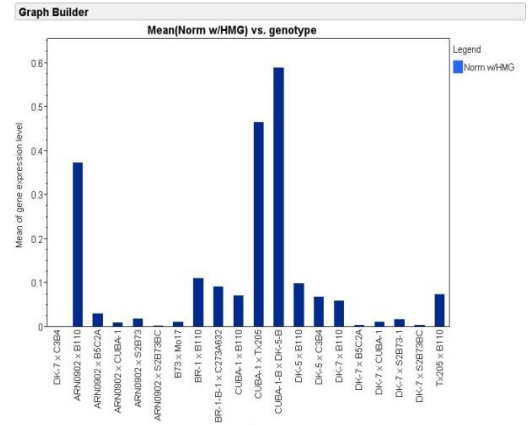
Fig. 5 (Continued)



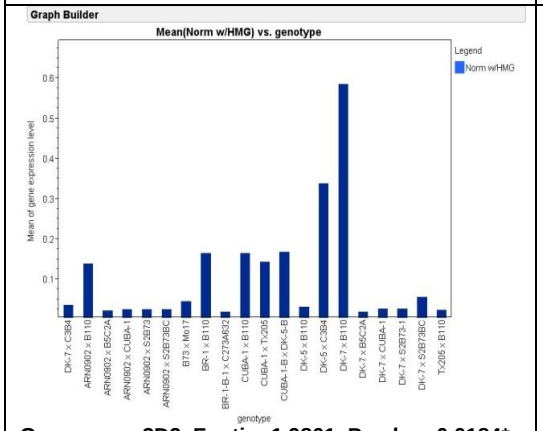
Gene name:2U2, F ratio=5.1840, P-value<.0001*



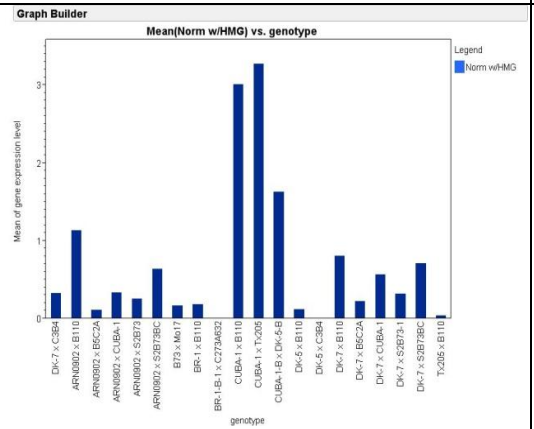
Gene name:2U3, F ratio=2.5766, P-value=0.0019*



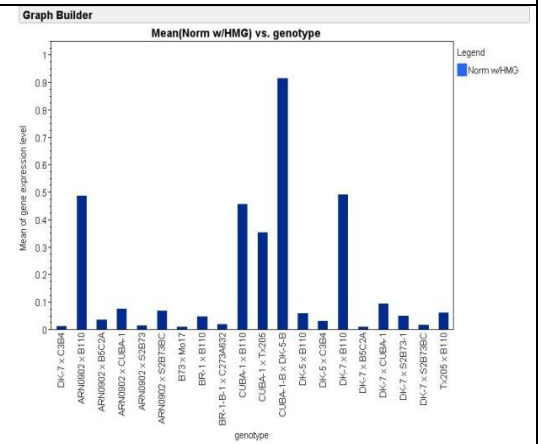
Gene name:2U4, F ratio=5.6208, P-value<.0001*



Gene name:3D2, F ratio=1.9861, P-value=0.0184*

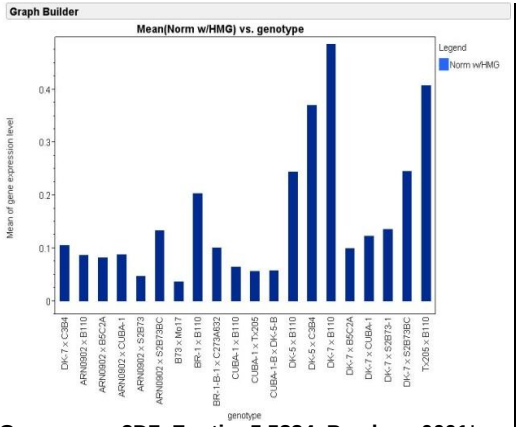


Gene name:3D3, F ratio=7.0556, P-value<.0001*

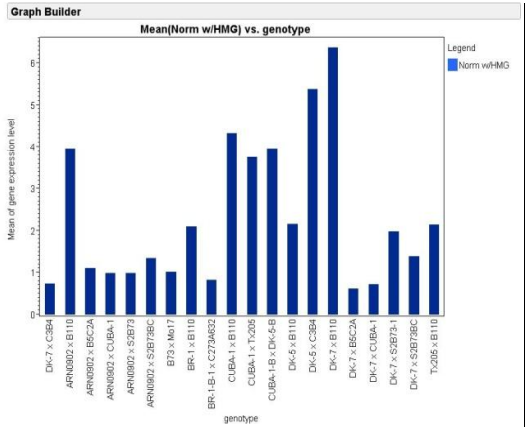


Gene name:3D4, F ratio=2.7504, P-value=0.0009*

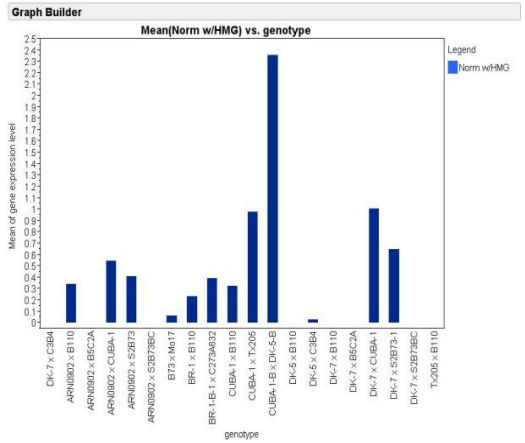
Fig. 5 (Continued)



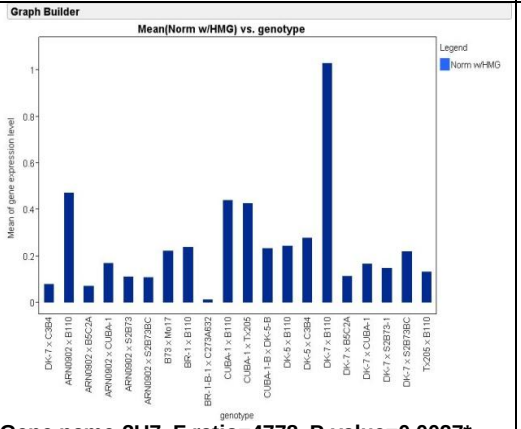
Gene name:3D7, F ratio=5.5884, P-value<.0001*



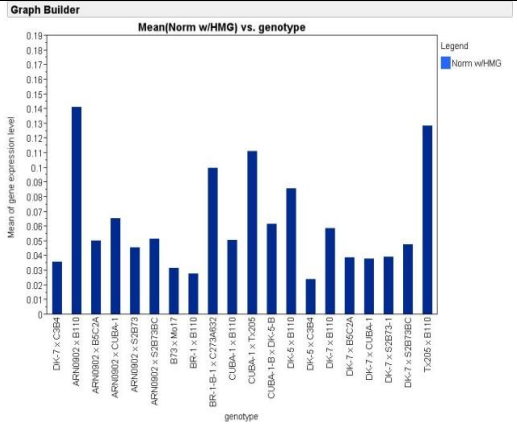
Gene name:3D9, F ratio=1.8899, P-value=0.0264*



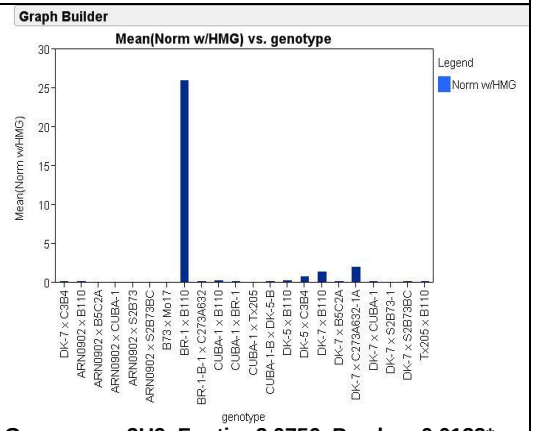
Gene name:3U2, F ratio=23.1487, P-value<.0001*



Gene name:3U7, F ratio=4778, P-value=0.0027*

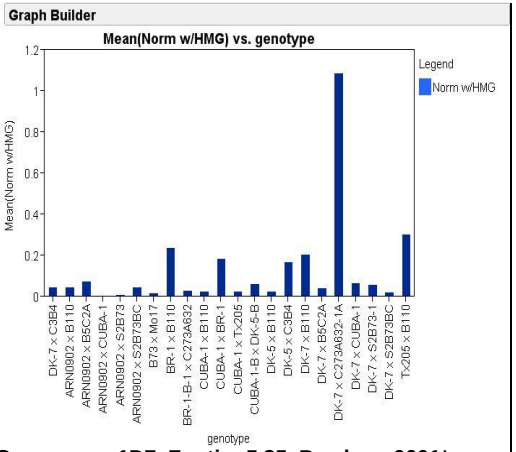


Gene name:3U8, F ratio=2.5531, P-value=0.0021*

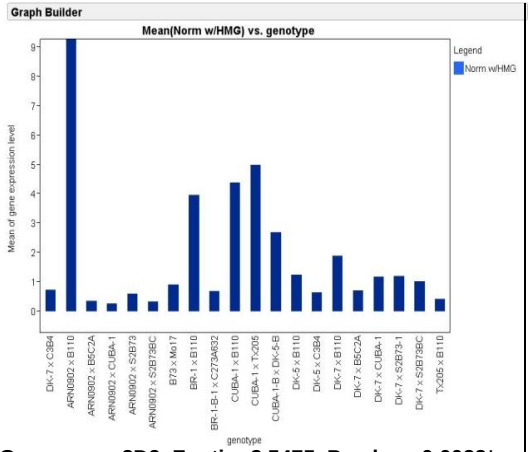


Gene name:3U9, F ratio=2.0756, P-value=0.0133*

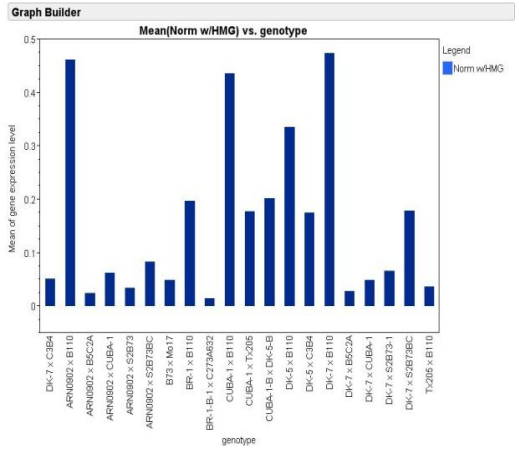
Fig. 5 (Continued)



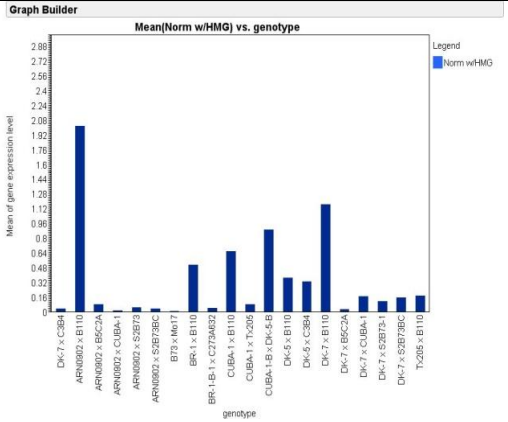
Gene name:1D7, F ratio=5.85, P-value<.0001*



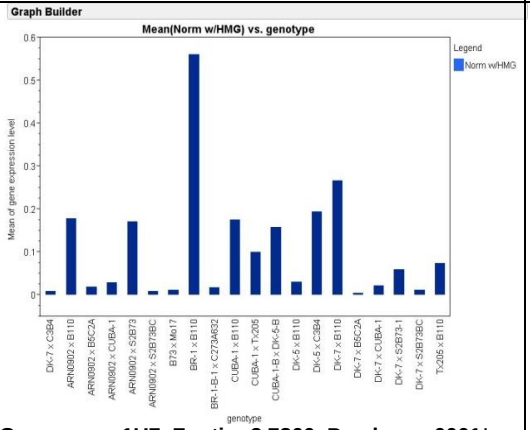
Gene name:2D9, F ratio=2.5475, P-value =0.0022*



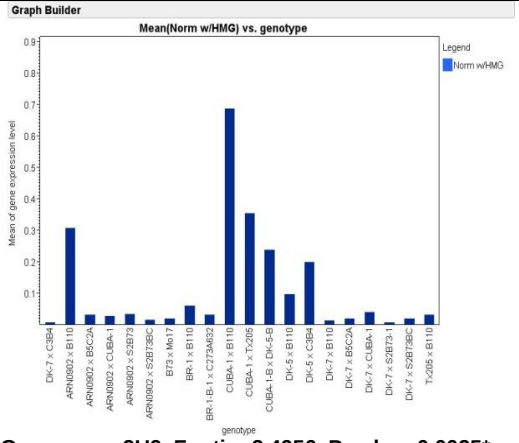
Gene name:2D10, F ratio=2.2920, P-value=0.0056*



Gene name:1U5, F ratio=2.9758, P-value=0.0004*

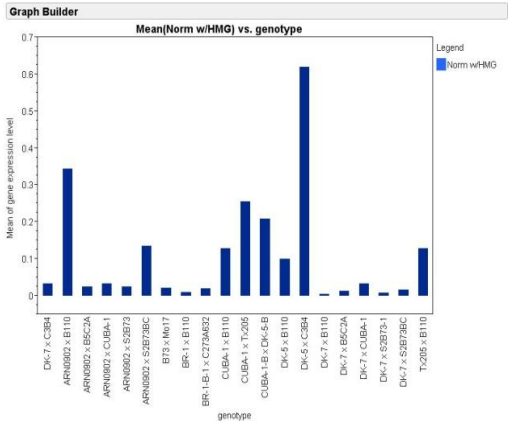


Gene name:1U7, F ratio=3.7860, P-value=<.0001*

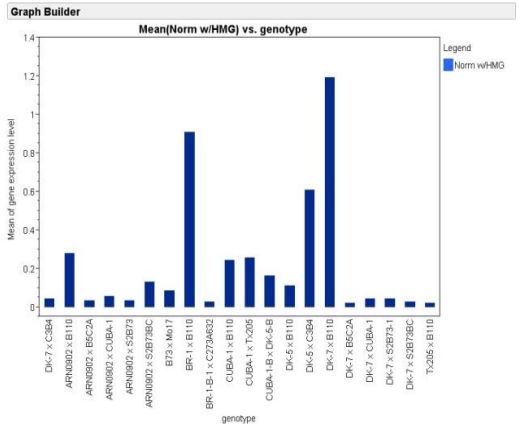


Gene name:2U9, F ratio=2.4956, P-value=0.0025*

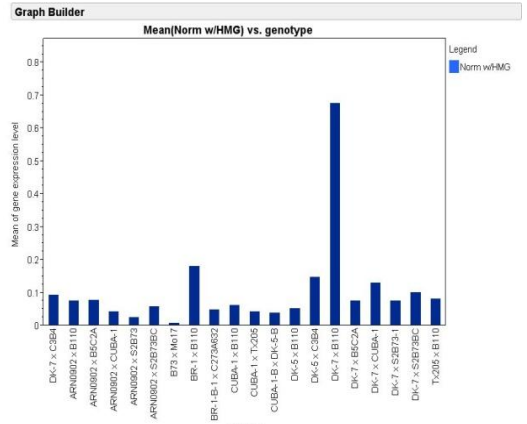
Fig. 5 (Continued)



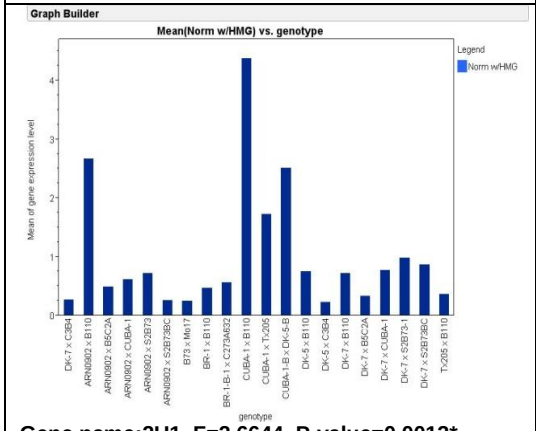
Gene name:2D4, F ratio=2.0123, P-value=0.0174*



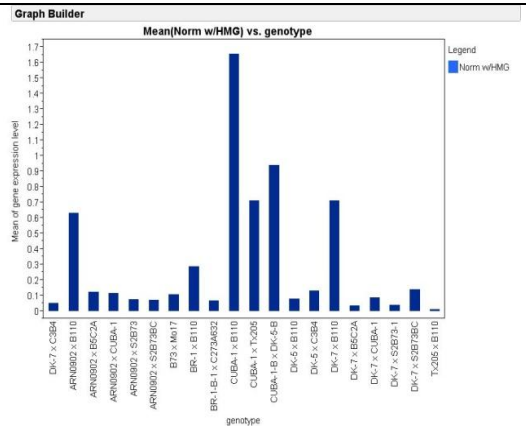
Gene name:2D5, F ratio=2.6612, P-value=0.0013*



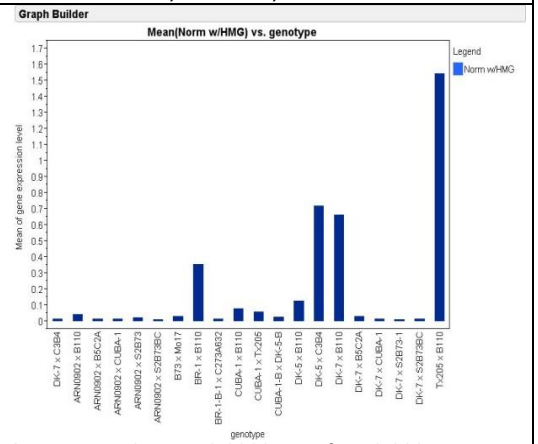
Gene name:3D5, F=3.1468, P-value=0.0002*



Gene name:2U1, F=2.6644, P-value=0.0013*

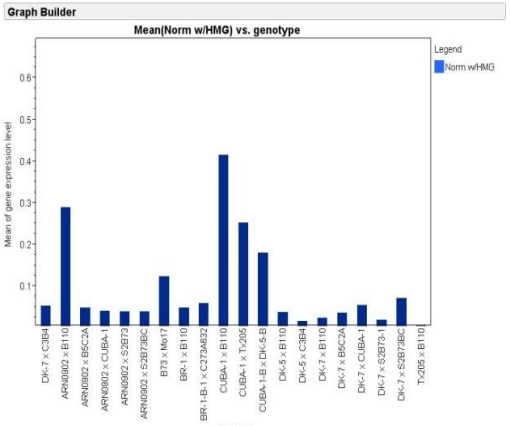


Gene name:3U10, F=1.9325, P-value=0.0225*

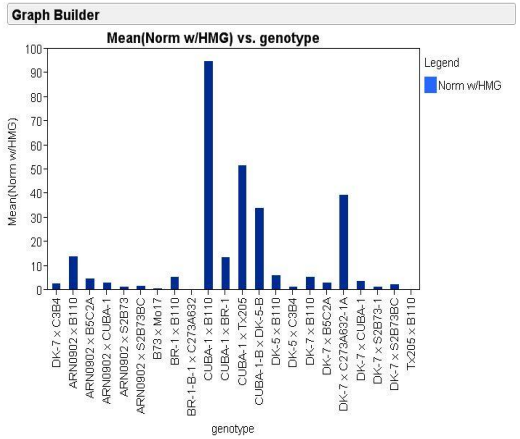


Gene name:3U4, F=2.8300, P-value=0.0007*

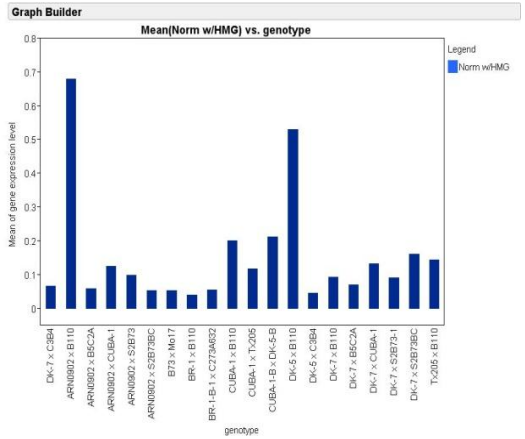
Fig. 5 (Continued)



Gene name:3D1, F=4.0334, P-value<.0001*

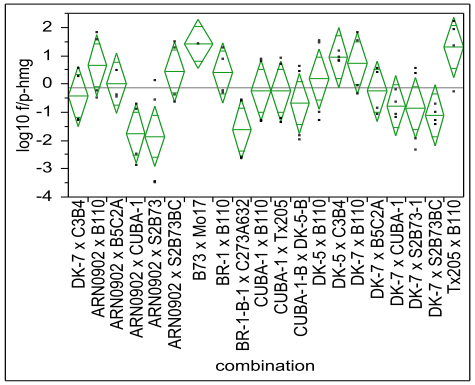


Gene name:3D6, F=4.2840, P-value<.0001*

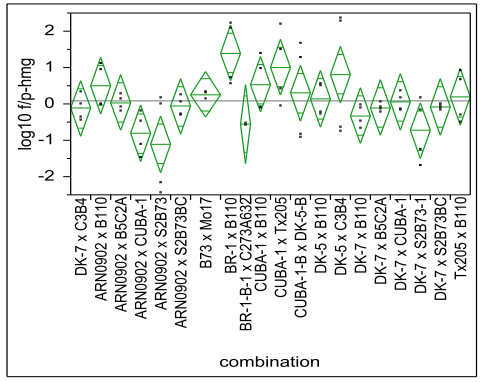


Gene name:3U5, F=2.4117, P-value=0.0036*

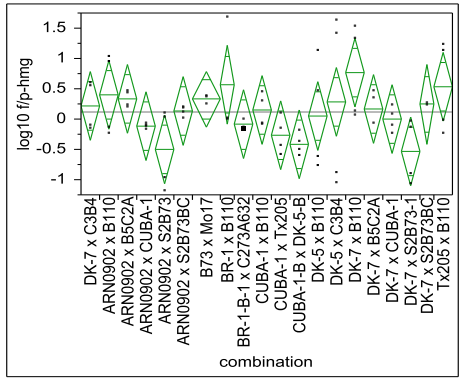
Fig. 5 (Continued)



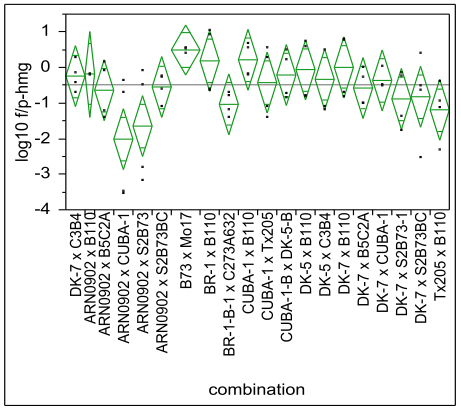
Gene name: 1D3, F Ratio=3.4635, P-value<.0001*



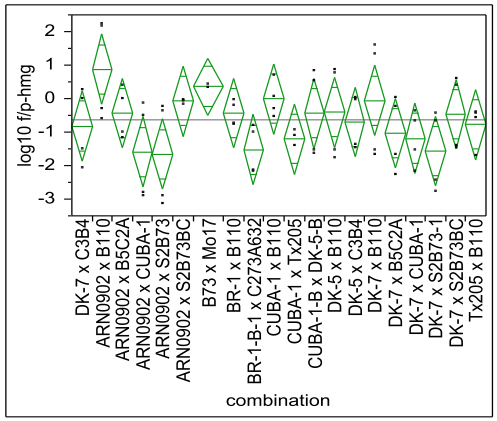
Gene name: 1D4, F Ratio=2.4864, P-value=0.0039*



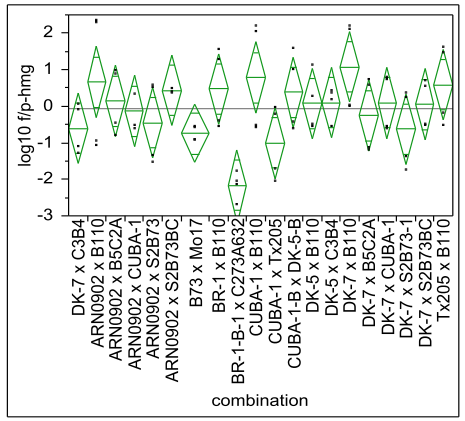
Gene name: 3U9, F Ratio=1.5362, P-value=0.1057



Gene name: 1D9, F Ratio=2.2549, P-value=0.0089*

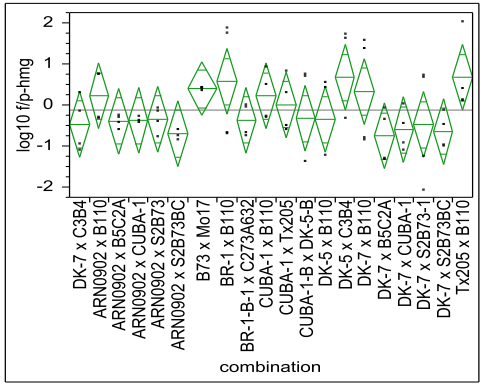


Gene name: 1D10, F Ratio=1.8349, P-value=0.0382*

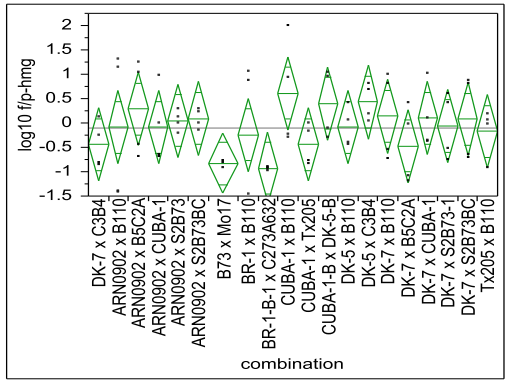


Gene name: 1U2, F Ratio=2.3732, P-value=0.0056*

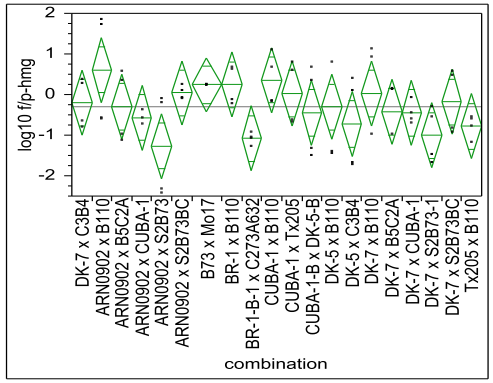
Fig. 6 Variation of eMPH of 39genes among 20 combinations of inbreds and F1 Hybrids. The Y axis represents $\log_{10}$ transformed eMPH and X axis stands for different combinations between inbreds and hybrids. Genes with a mean of eMPH below zero represents down-regulation in the hybrids while those with a mean of eMPH above zero are up-regulation in the hybrids, relative to the mean of their parents.



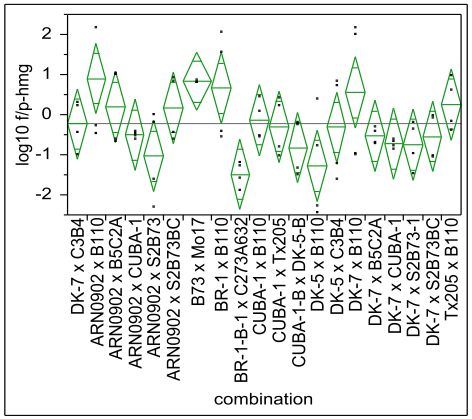
Gene name: 1U7, F Ratio=1.4522, P-value=0.1365



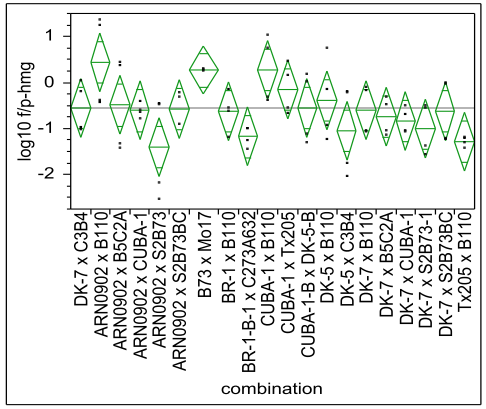
Gene name: 1U9, F Ratio=1.2157, P-value=0.2750



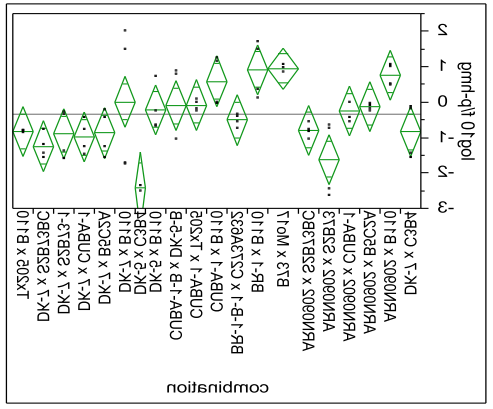
Gene name: 2D1, F Ratio=1.6162, P-value=0.0803



Gene name: 2D5, F Ratio=2.5206, P-value=0.0033*

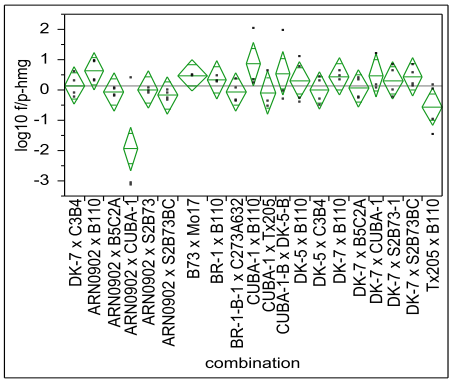


ene name: 2D6, F Ratio=2.6398, P-value=0.0021*

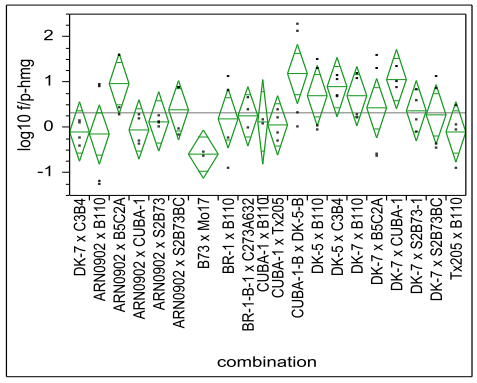


Gene name: 2D7, F Ratio=5.3940, P-value<.0001*

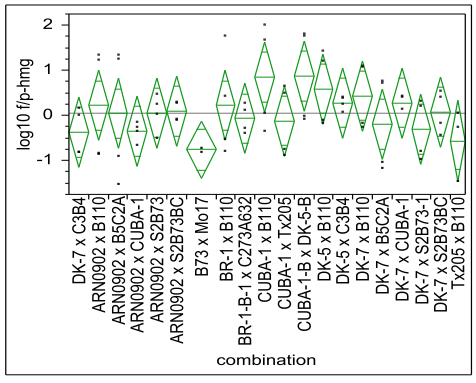
Fig. 6 (Continued)



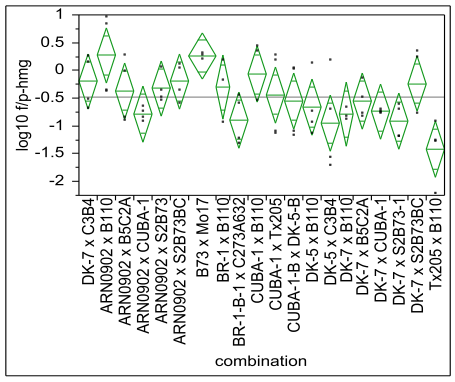
Gene name: 2U1, F Ratio=2.8912, P-value=0.0010*



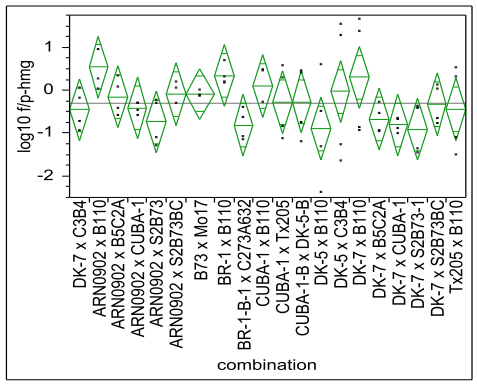
Gene name: 2U2, F Ratio=2.1450, P-value=0.0132*



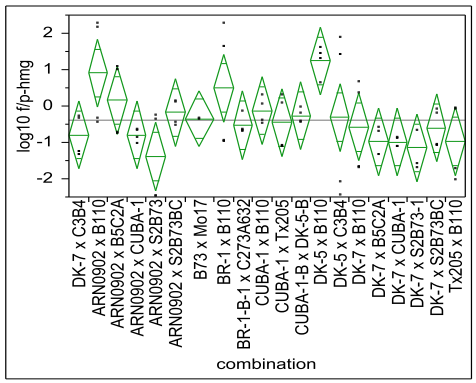
Gene name: 2U3, F Ratio=1.3040, P-value=0.2146



Gene name: 3D1, F Ratio=2.9869, P-value=0.0007*

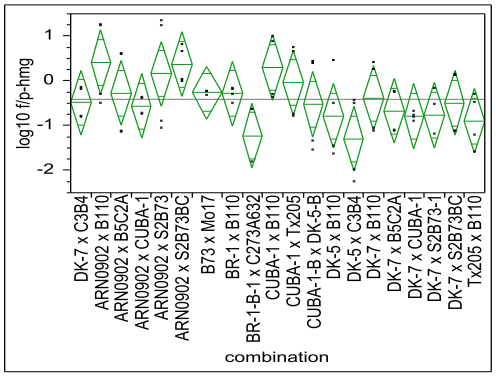


Gene name: 3D2, F Ratio=1.4039, P-value=0.1585

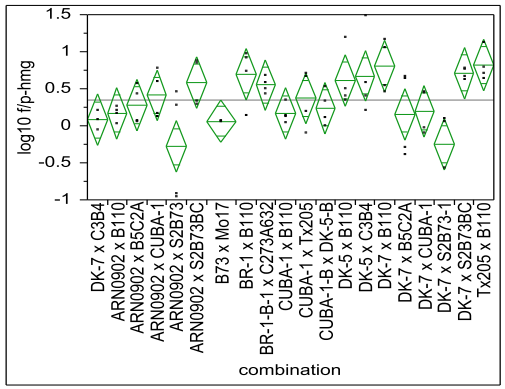


Gene name: 3D3, F Ratio=2.0892, P-value=0.0156*

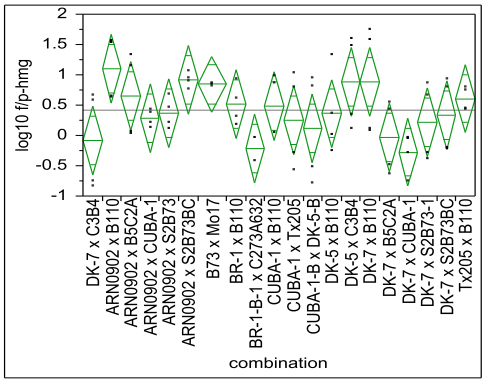
Fig. 6 (Continued)



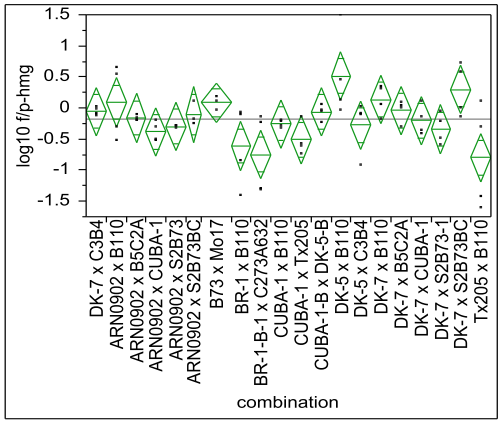
Gene name: 3D6, F Ratio=1.8114, P-value=0.0414*



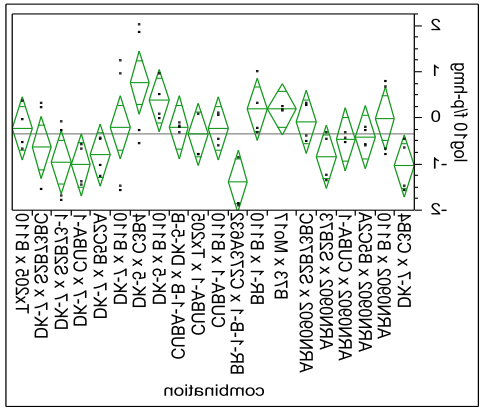
Gene name: 3D7, F Ratio=3.6151, P-value<.0001*



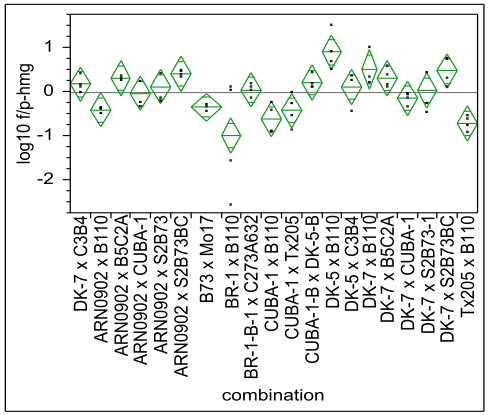
Gene name: 3D9, F Ratio=2.0197, P-value=0.0199*



Gene name: 3U5, F Ratio=2.9522, P-value=0.0007*

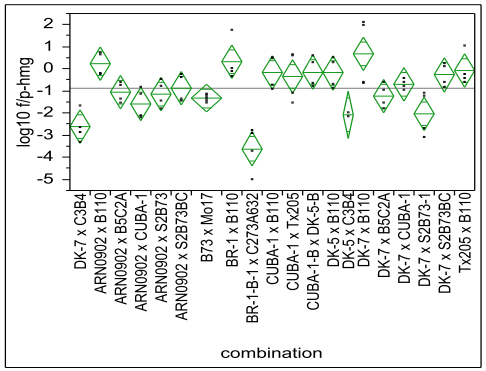


Gene name: 3U7, F Ratio=2.5763, P-value=0.0027*

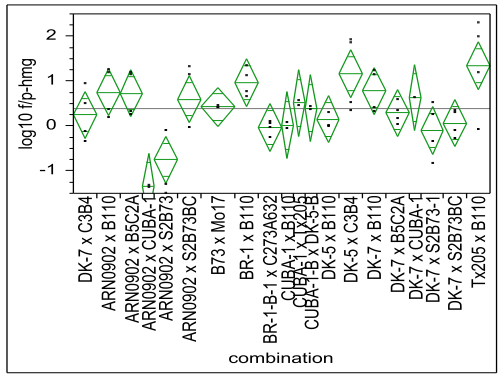


Gene name: 3U8, F Ratio=5.6292, P-value<.0001*

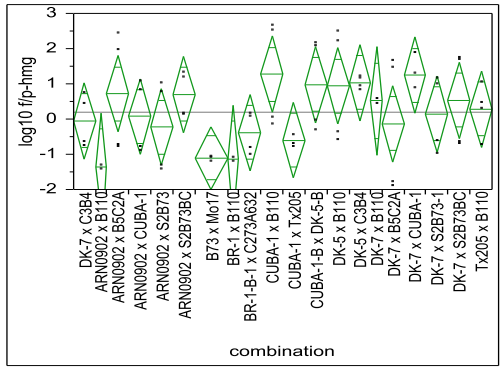
Fig. 6 (Continued)



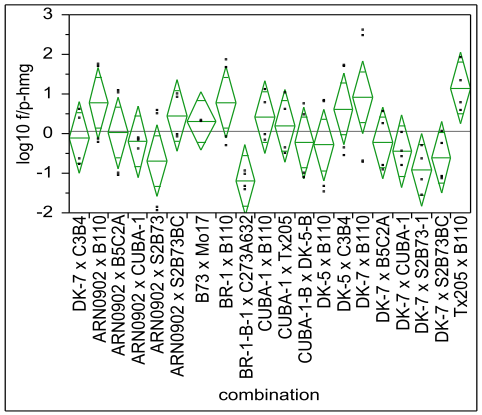
Gene name: 1D6, F Ratio=7.8665, P-value<.0001*



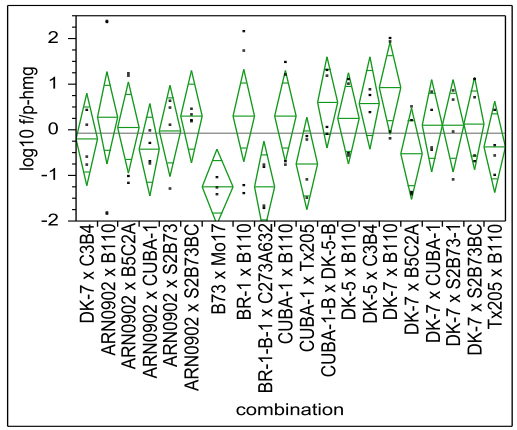
Gene name: 1D7, Ratio=4.4588, P-value<.0001*



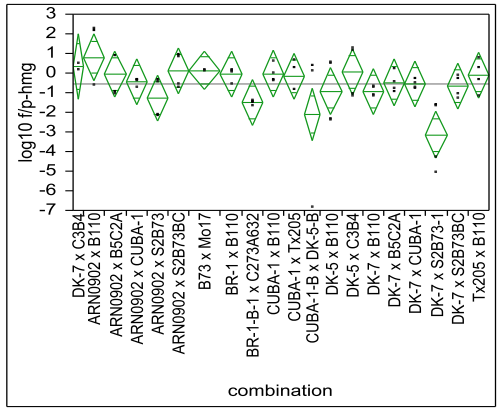
Gene name: 1U4, F Ratio=1.9547, P-value=0.0273*



Gene name: 2D3, F Ratio=1.9863, P-value=0.0224*

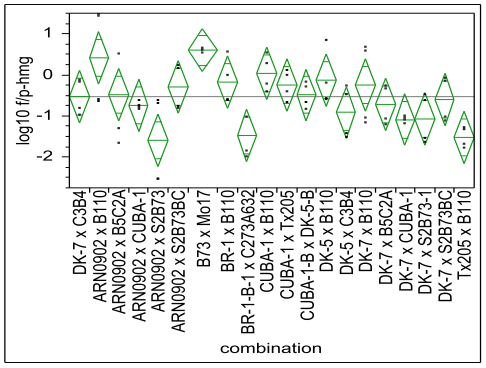


Gene name: 1U5, F Ratio=1.4453,P-value=0.1394

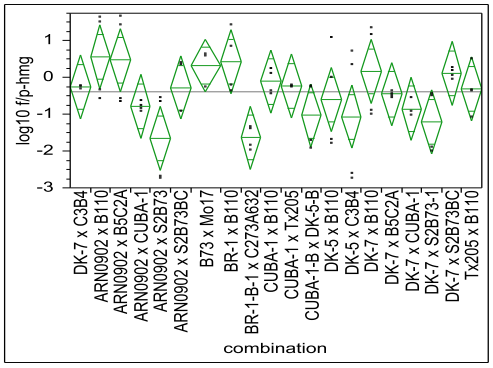


Gene name: 2D4, F Ratio=2.3078, P-value=0.0076*

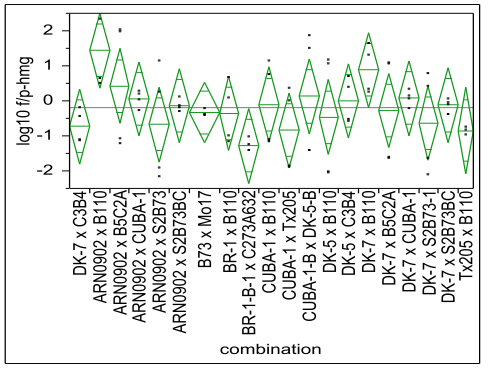
Fig. 6 (Continued)



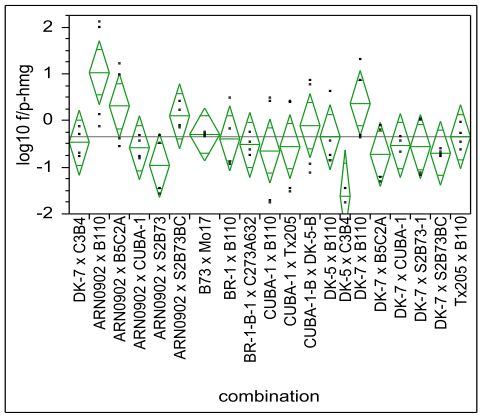
Gene name: 2D9, F Ratio=3.9004, P-value<.0001*



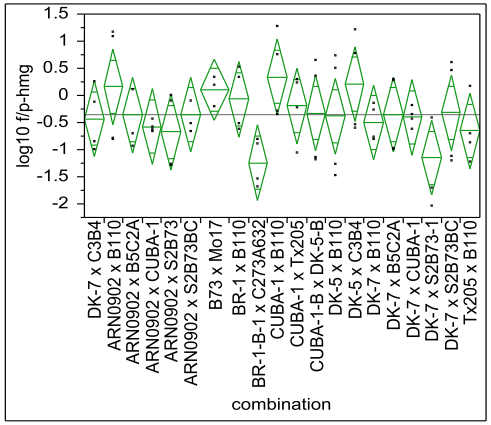
Gene name: 2D10, F Ratio=2.5006, P-value=0.0035*



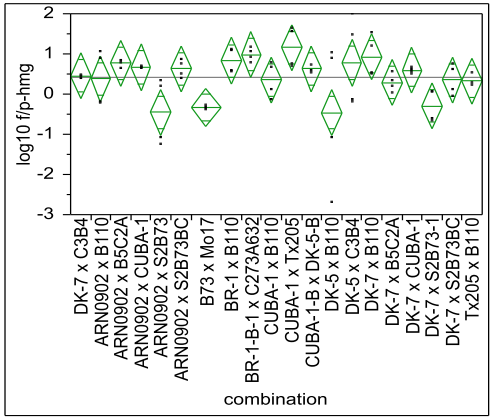
Gene name: 2U4, F Ratio=1.3543, P-value=0.1850



Gene name: 3D4, F Ratio=2.1580, P-value=0.0126*

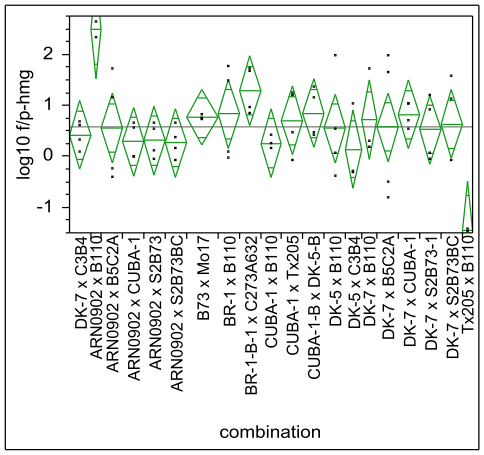


Gene name: 2U9, F Ratio=1.3804, P-value=0.1704

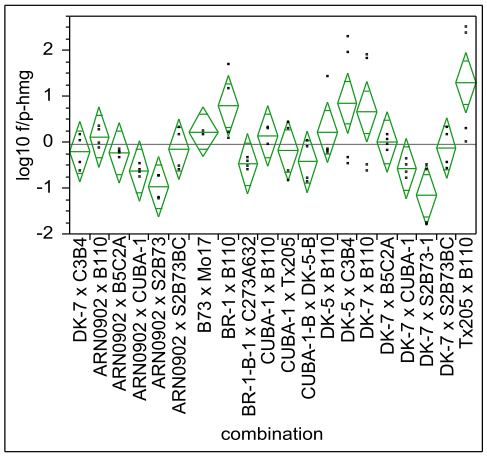


Gene name: 3D5,F Ratio=2.9938, P-value=0.0006*

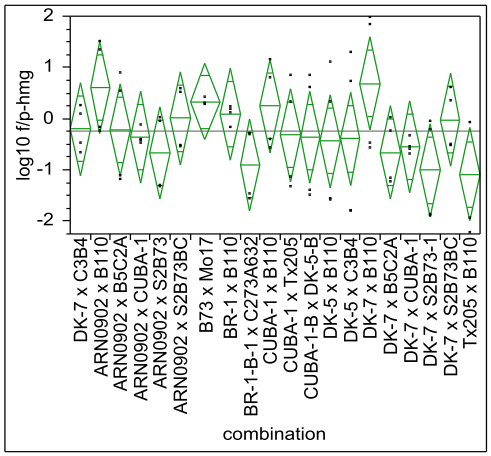
Fig. 6 (Continued)



Gene name: 3u2, F Ratio=2.4002, P-value=0.0059*



Gene name: 3u4, F Ratio=3.4626, P-value=0.0001*



Gene name: 3u10, F Ratio=1.2438, P-value=0.2543

Fig. 6 (Continued)

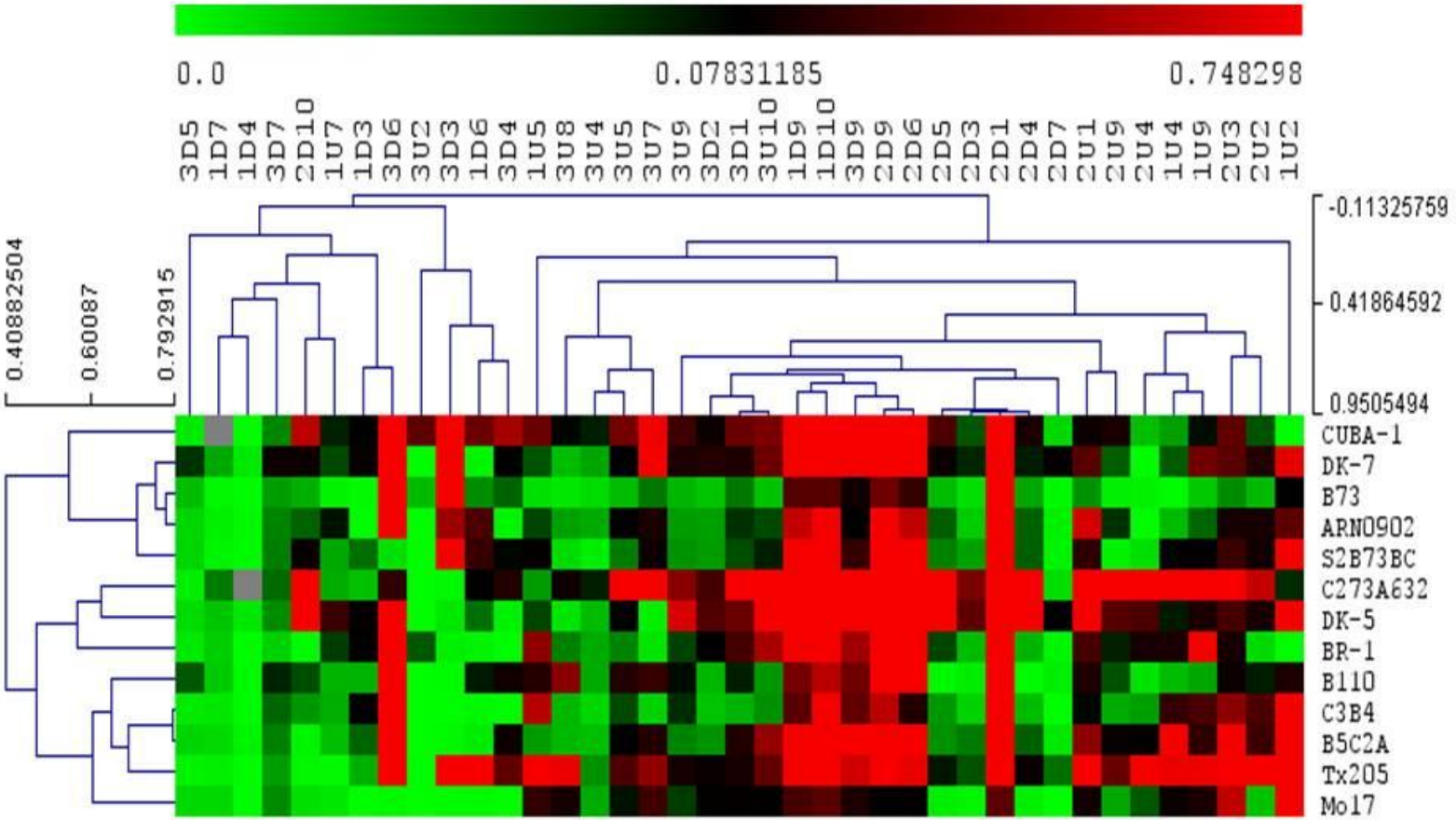


Fig. 7 Hierarchical clustering heat map of inbred lines by Spearman's ρ(P-value≤0.05)

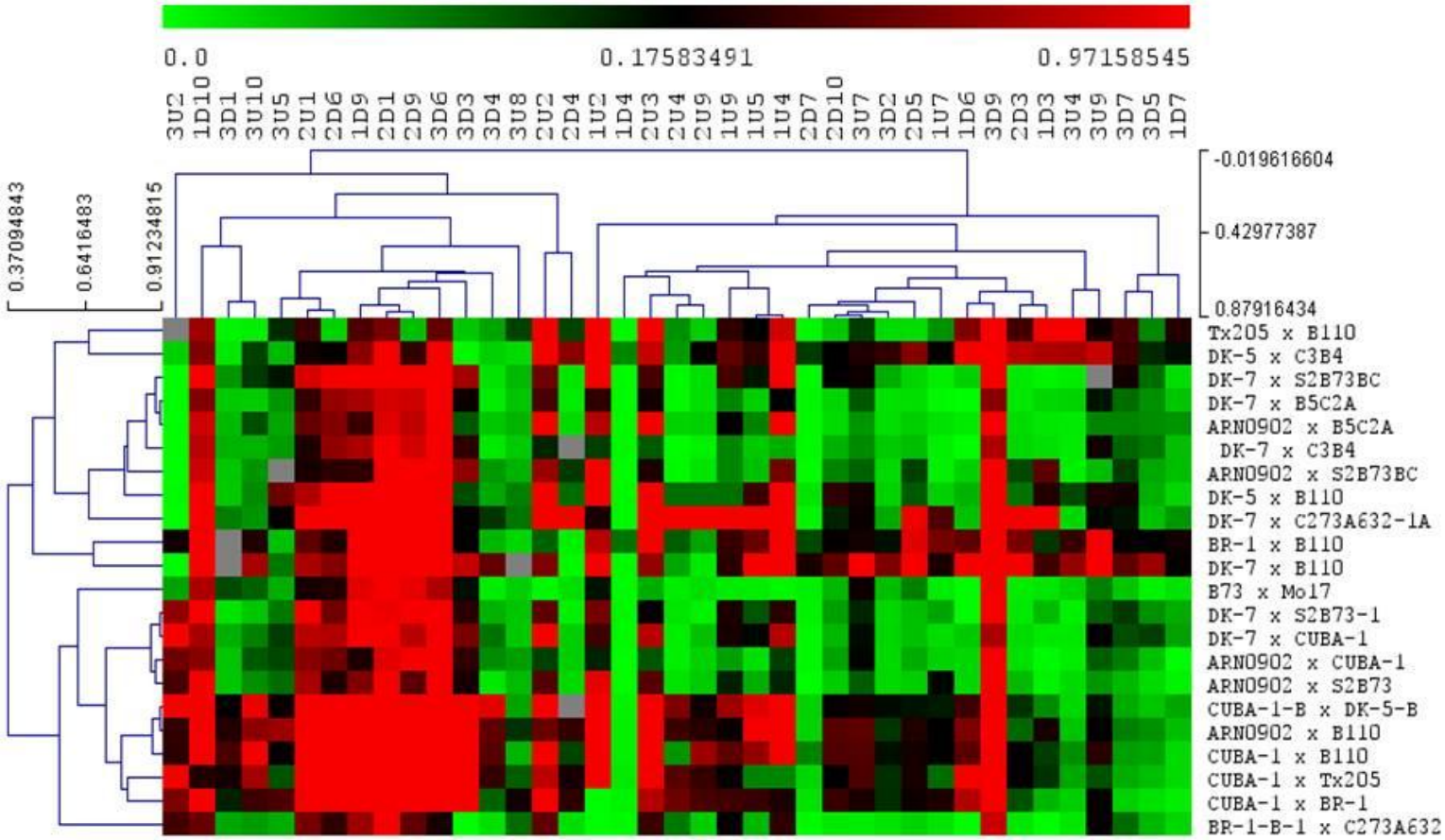


Fig. 8 Hierarchical clustering heat map of hybrid lines by Spearman's ρ(P-value≤0.05)

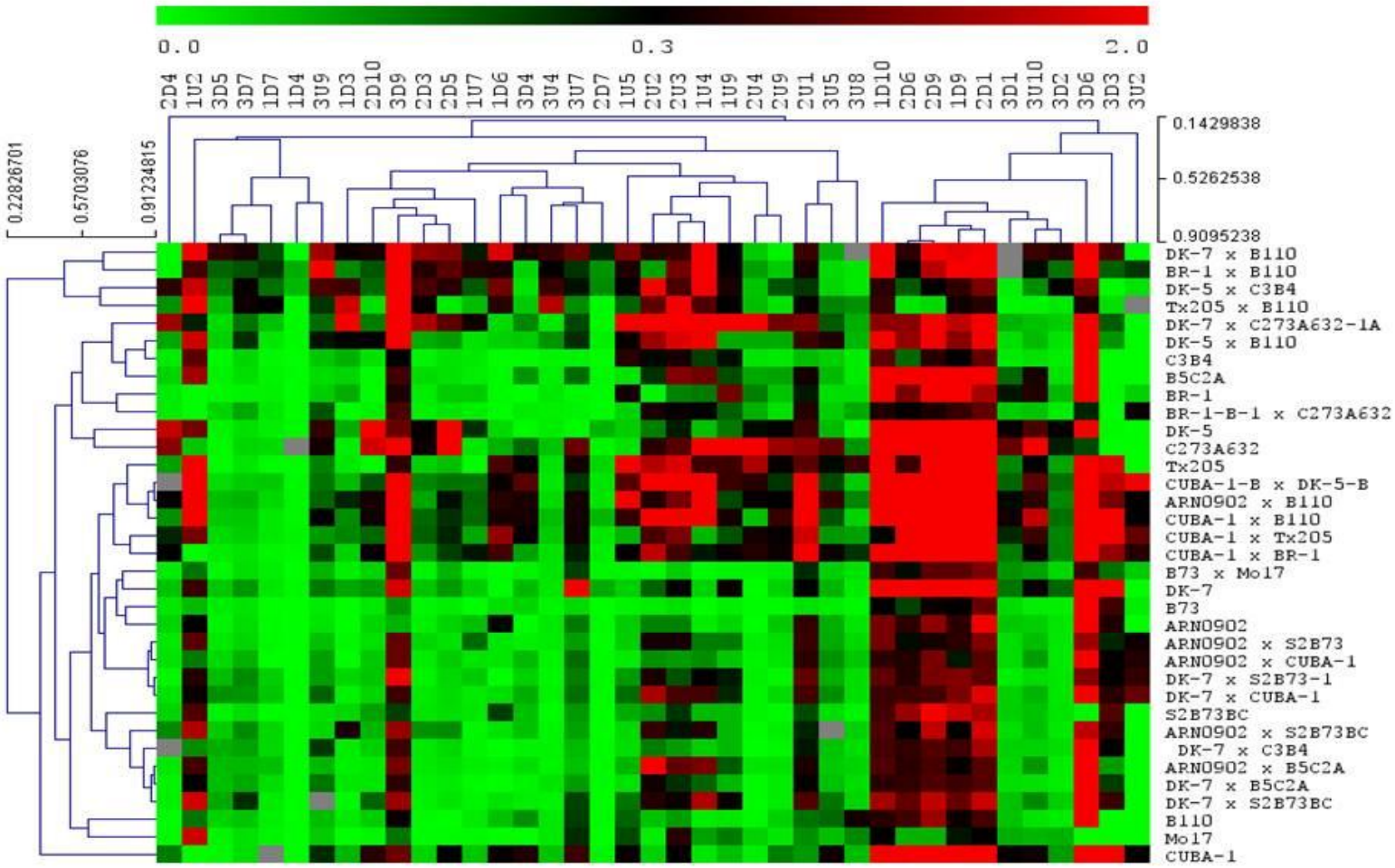


Fig. 9 Hierarchical clustering heat map of hybrids and inbreds by Spearman's ρ (P-value≤0.05)

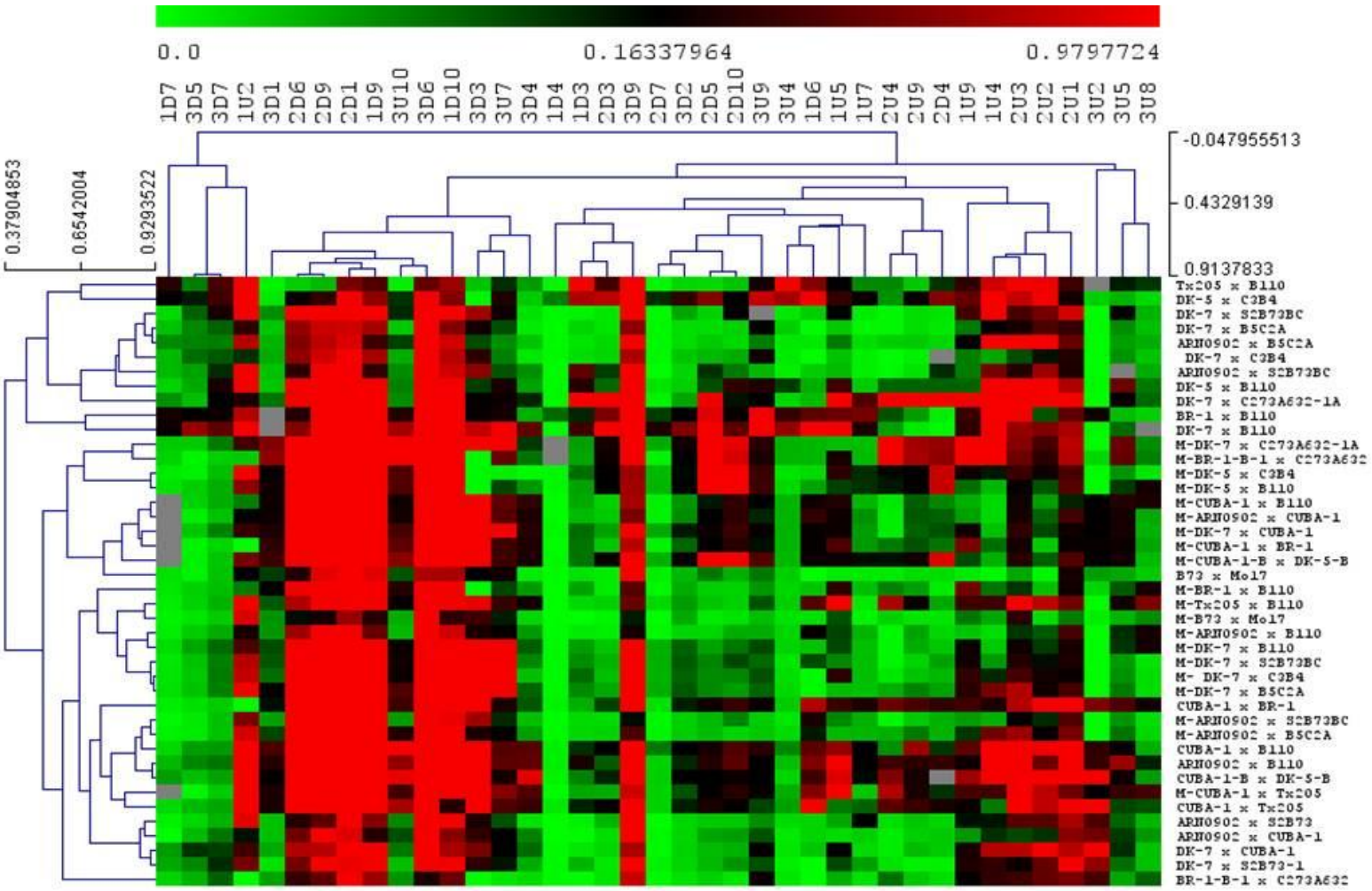


Fig. 10 Hierarchical clustering heat map of hybrids and Mid-inbred-parents by Spearman's ρ (P-value≤0.05)

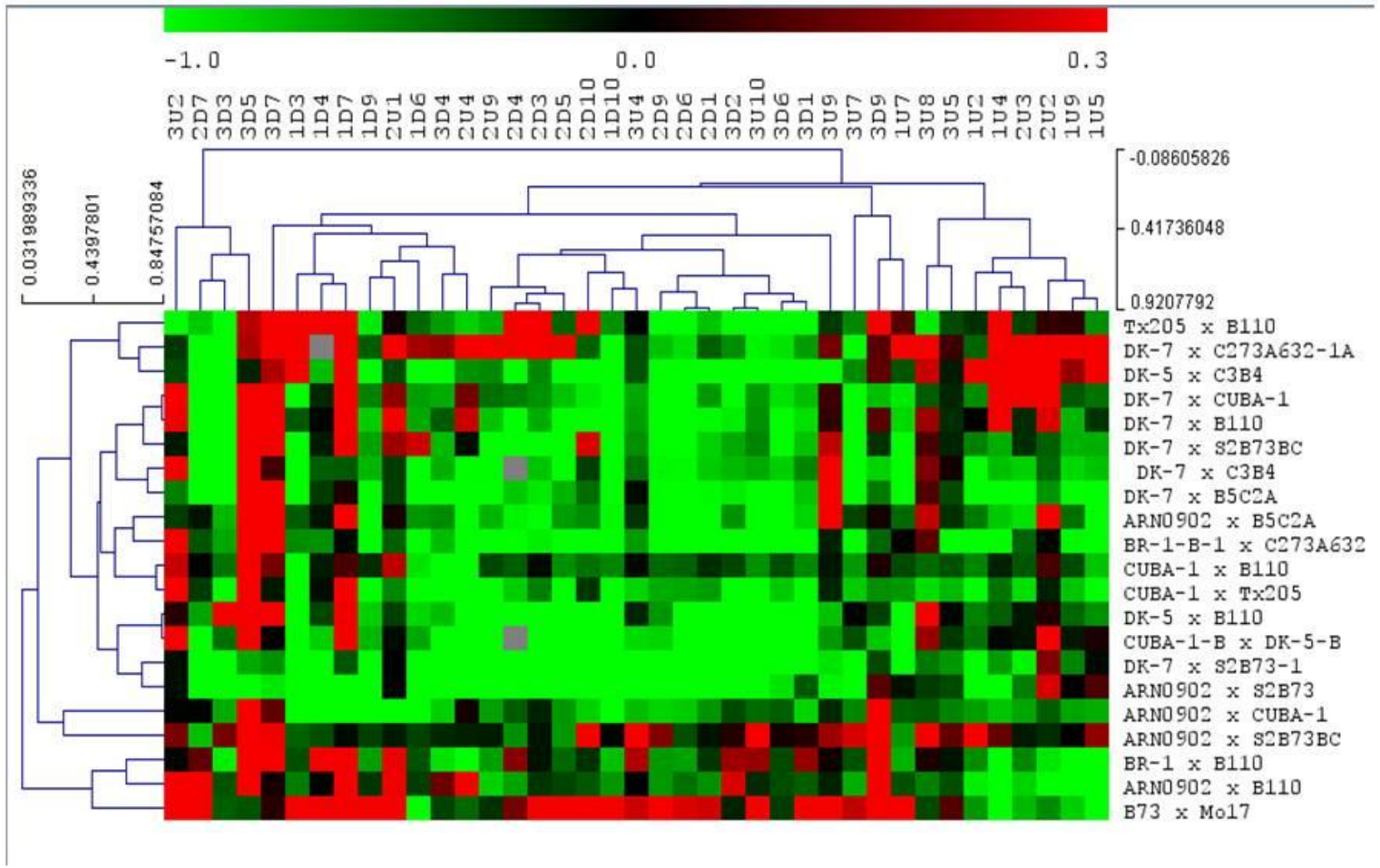
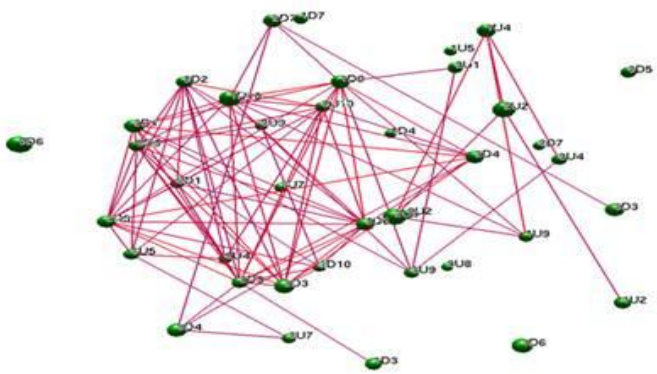
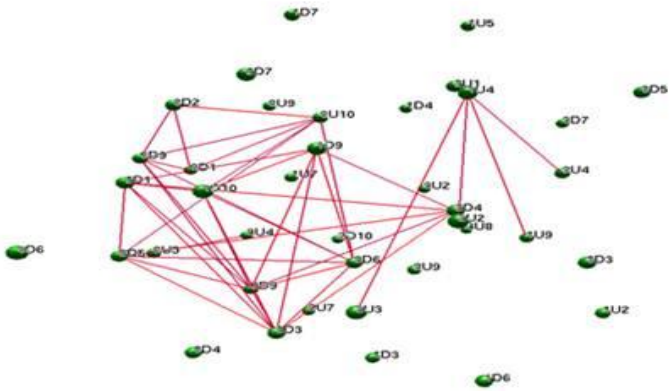


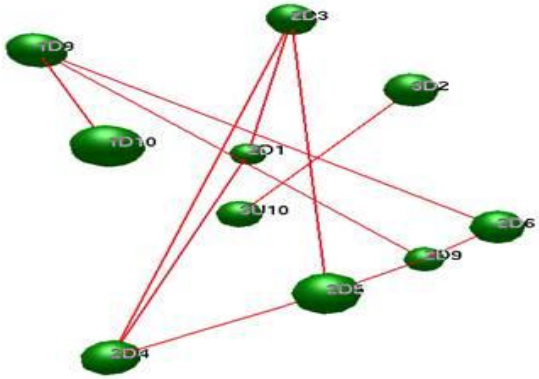
Fig. 11 Hierarchical clustering heat map of expression MPH by Spearman's ρ (P-value≤0.05)



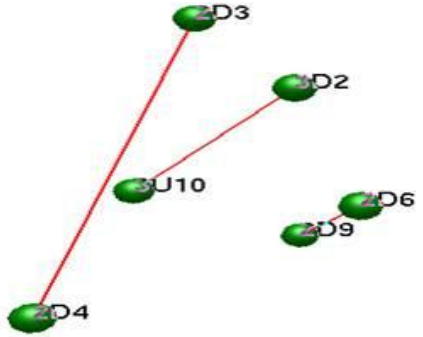
A. Spearman's ρ≥ 0.6, P-value≤0.0003



B. Spearman's ρ≥ 0.7. P-value≤0.0001

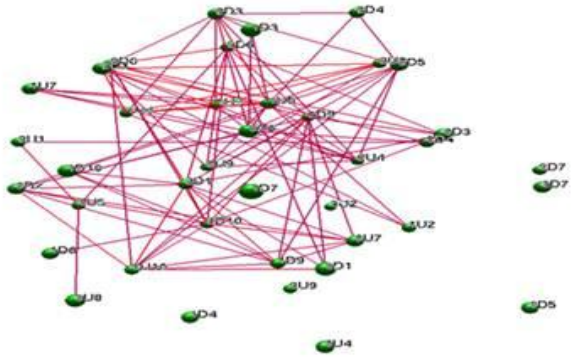


C. Spearman's ρ≥ 0.8. P-value≤0.0001

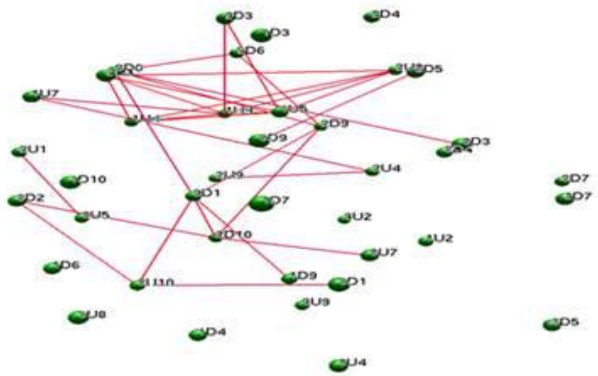


D. Spearman's ρ≥0.9, P-value≤0.0001

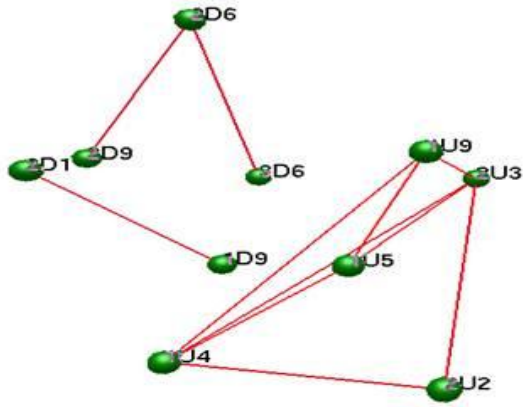
Fig. 12 Gene network in inbred lines under different level of Spearman's ρ



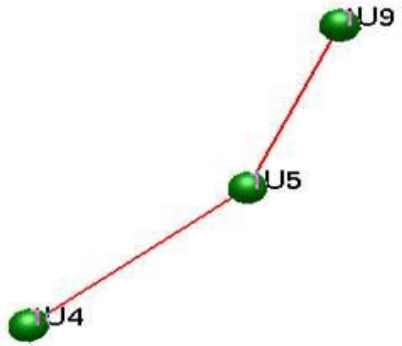
A Spearman's ρ≥0.6. P-value≤0.0003



B. Spearman's ρ≥0.7. P-value≤0.0001

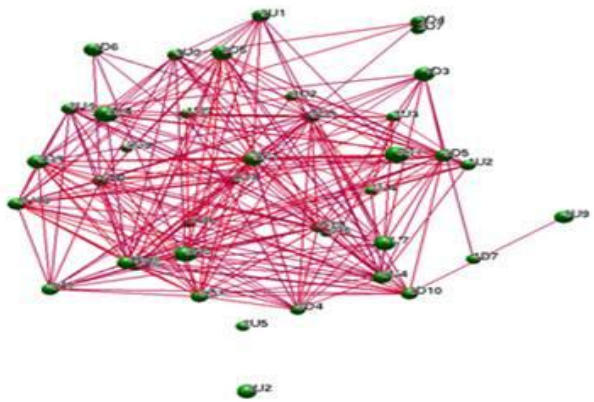


C Spearman's ρ≥0.8. P-value≤0.0001

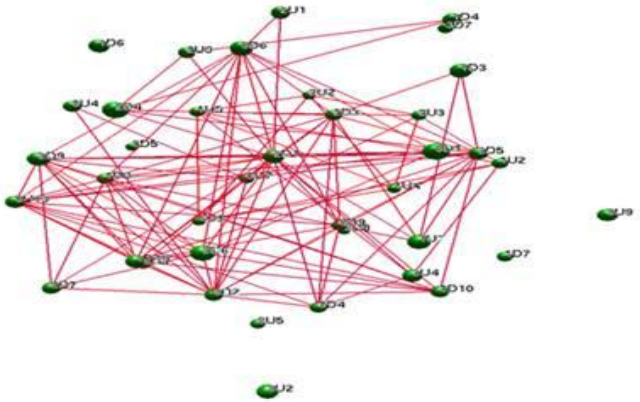


D. Spearman's ρ≥0.9. P-value≤0.0001

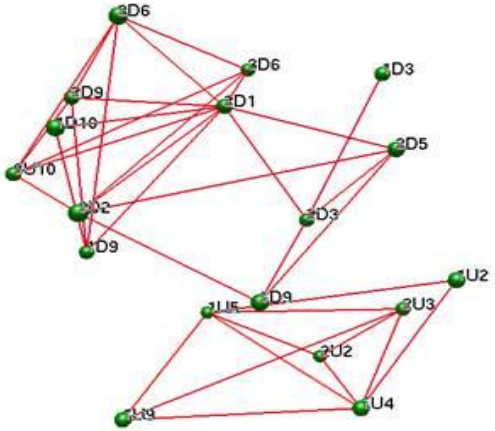
Fig. 13 Gene network in hybrid lines under different level of Spearman's ρ



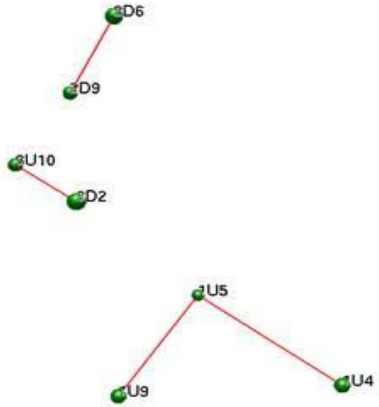
A. Spearman's ρ≥ 0.6, P-value≤0.0003



B. Spearman's ρ≥ 0.7. P-value≤0.0001



C. Spearman's ρ≥ 0.8. P-value≤0.0001



D. Spearman's ρ≥ 0.9, P-value≤0.0001

Fig. 14 Gene network in expression MPH under different level of Spearman's ρ

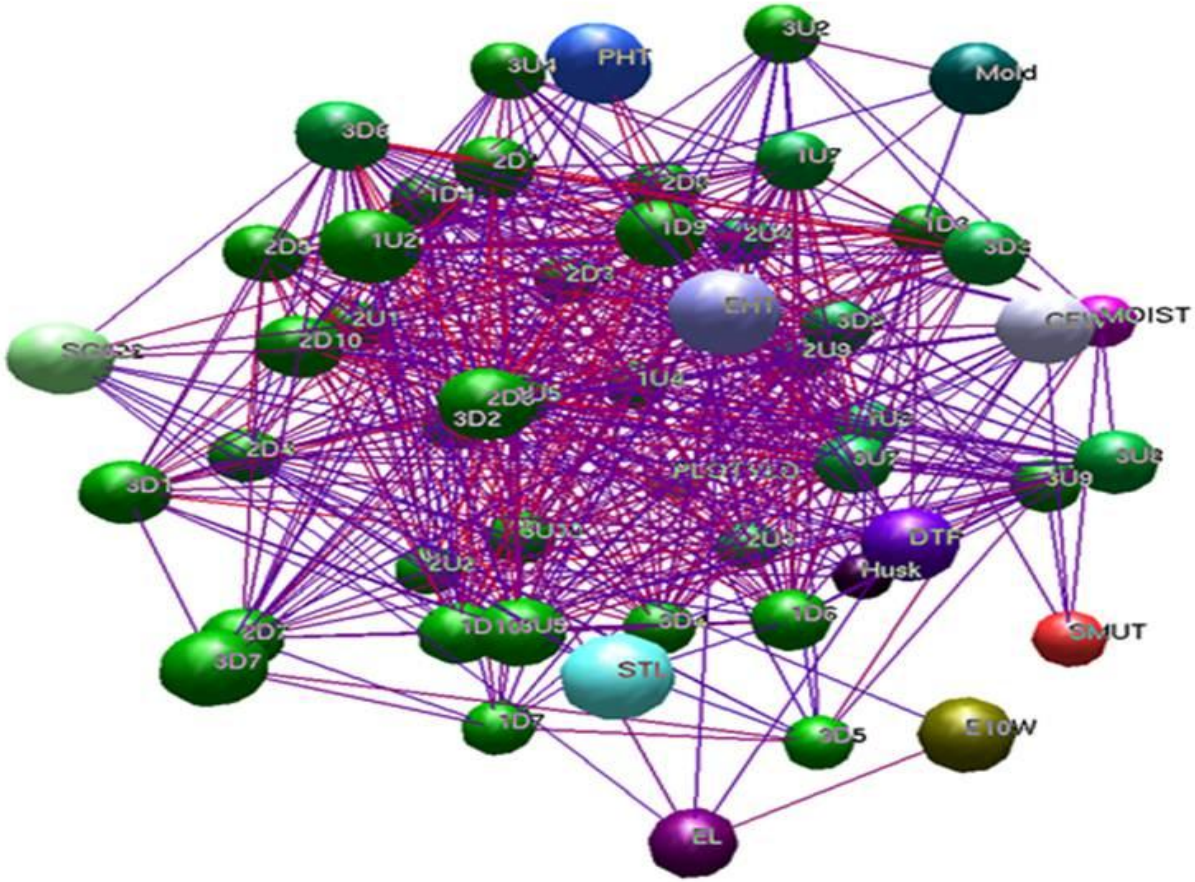
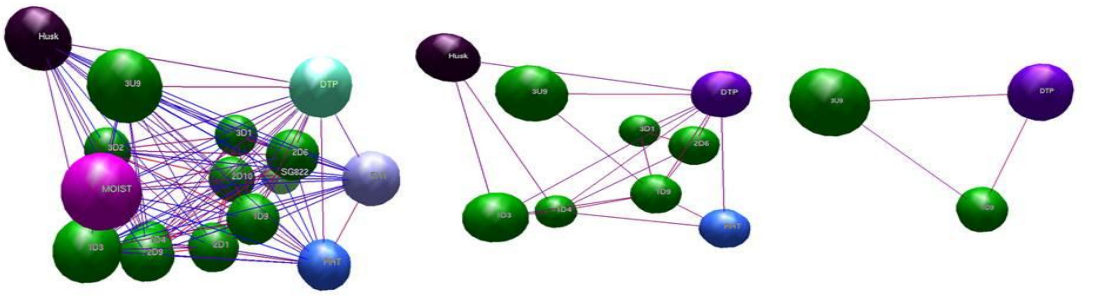


Fig. 15 Gene network of 12 traits. Spearman's ρ≥ 0.3. P-value≤0.03

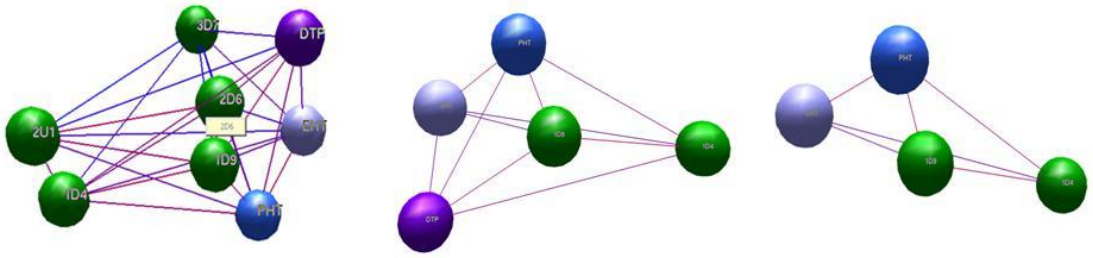


A. Spearman's ρ≥ 0.3 P-value≤0.03

B. Spearman's ρ≥ 0.4 P-value≤0.004

C. Spearman's ρ ≥ 0.5 P-value≤0.0001

Fig. 16 Gene network for days to pollen shedding(DTP)

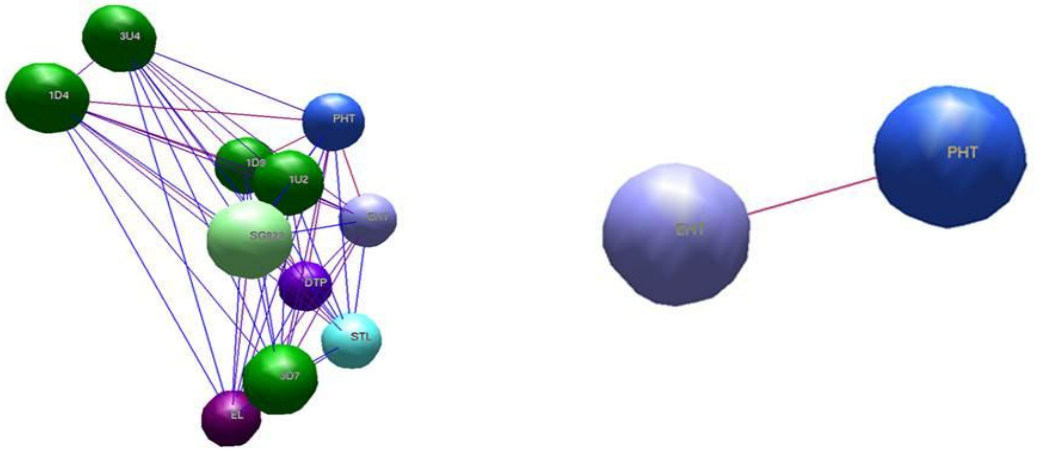


A. Spearman's ρ > 0.3 P-value<0.03

B. Spearman's ρ > 0.4 P-value<0.004

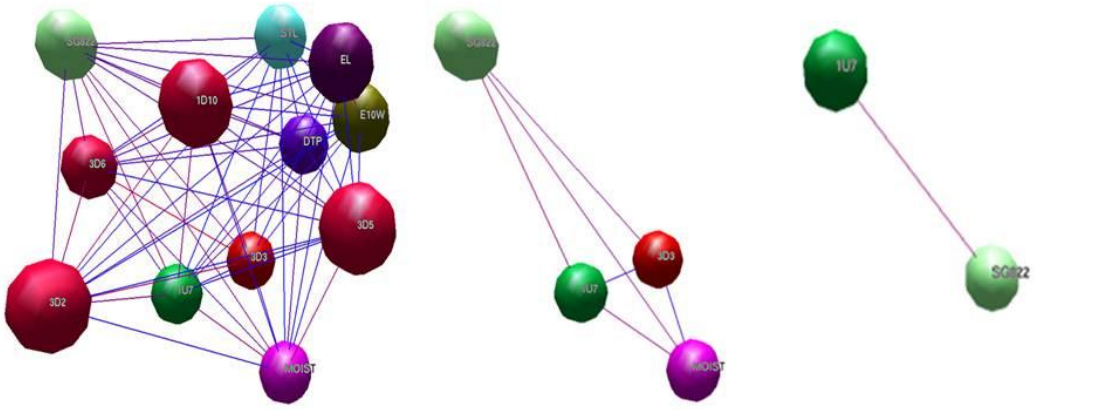
C. Spearman's ρ > 0.5 P-value<0.0001

Fig. 17 Gene network with trait "Plant height"(PHT).



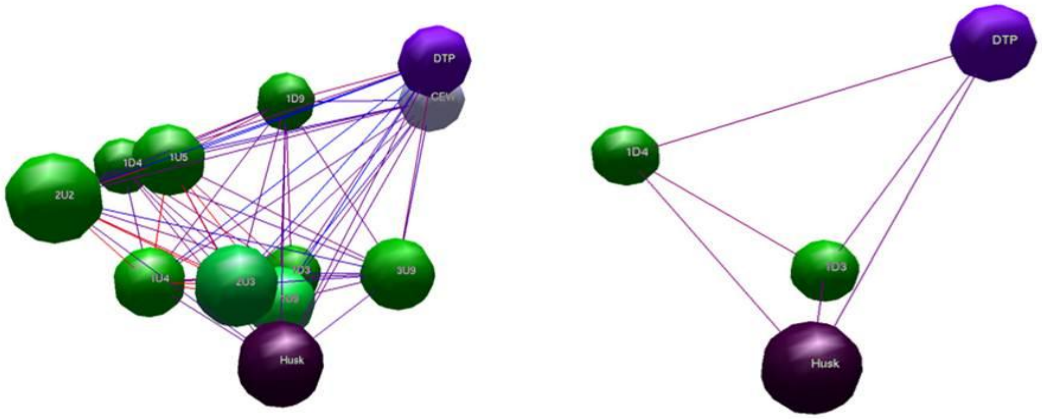
A. Spearman's ρ≥ 0.3. P-value≤0.03 **B. Spearman's ρ≥ 0.4. P-value≤0.004**

Fig. 18 Gene network for ear height(EHT).



A. Spearman's ρ≥ 0.3 P-value≤0.03 **B. Spearman's ρ≥ 0.4 P-value≤0.004** **C. Spearman's ρ ≥ 0.5 P-value≤0.0001**

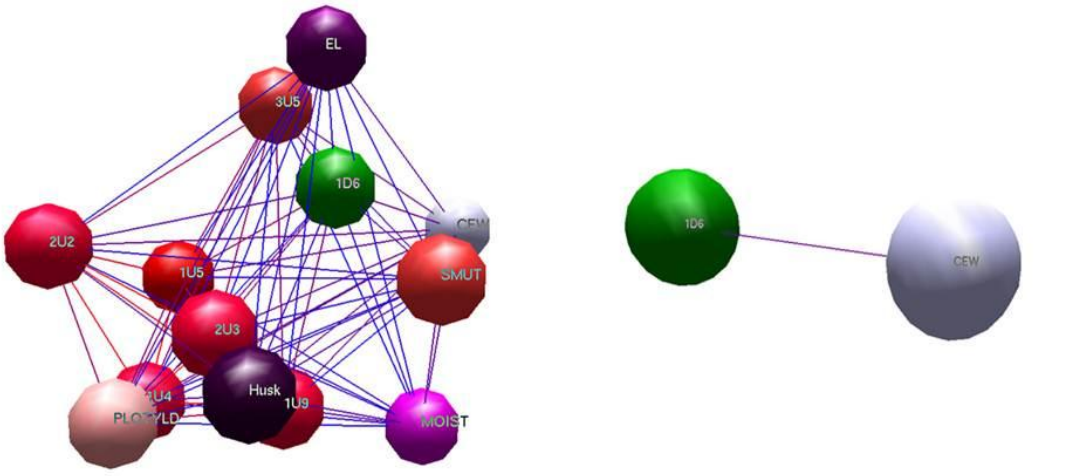
Fig. 19 Gene network for stay green rating on AUG 22(SG822).



A. Spearman's ρ≥ 0.3
P-value≤0.03

B. Spearman's ρ≥ 0.4
P-value≤0.004

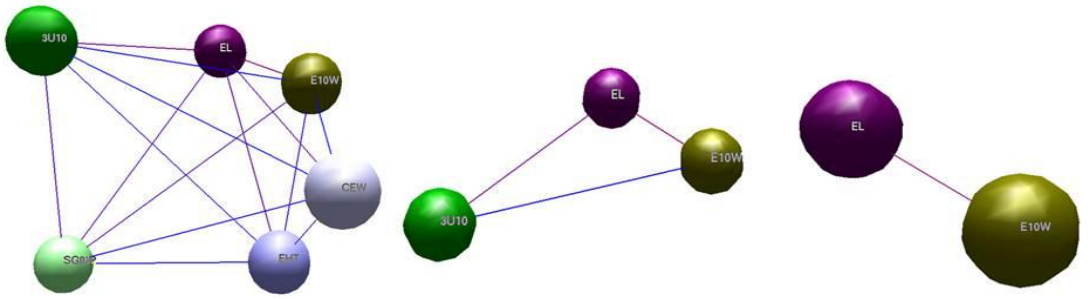
Fig. 20 Gene network for husk coverage rating(HUSK).



A. Spearman's ρ ≥ 0.3. P-value≤0.03

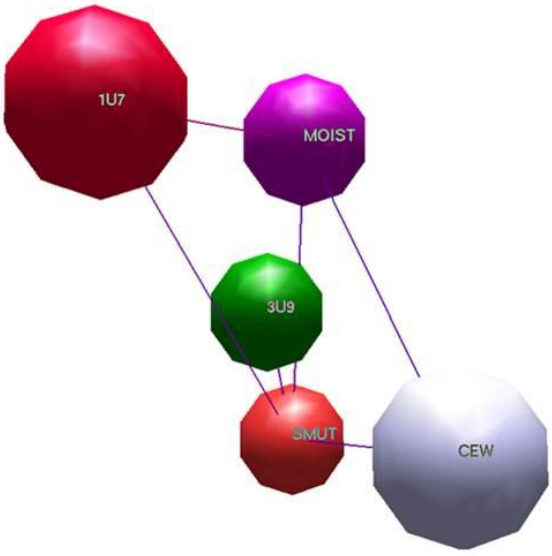
B. Spearman's ρ ≥ 0.4. P-value≤0.004

Fig. 21 Gene network for corn earworm feeding damage(CEW).



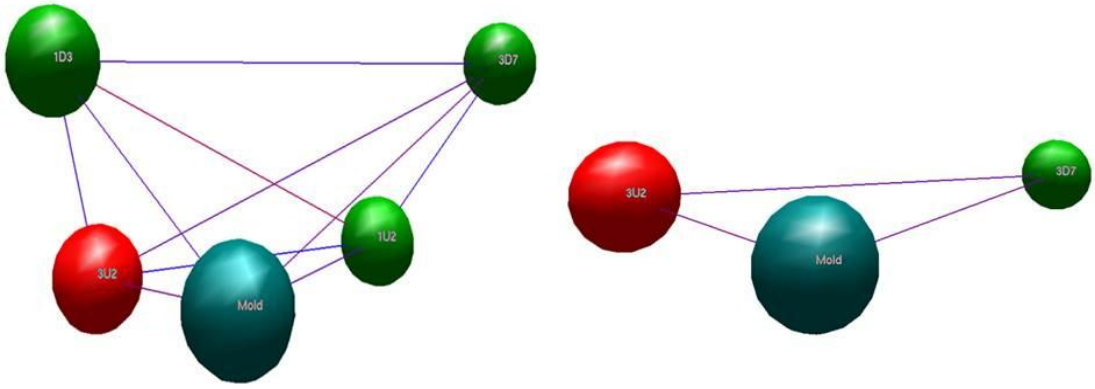
A. Spearman's ρ≥ 0.3 P-value≤0.03 **B. Spearman's ρ≥ 0.4 P-value≤0.004** **C. Spearman's ρ ≥ 0.5 P-value≤0.0001**

Fig. 22 Gene network for ear length (EL).



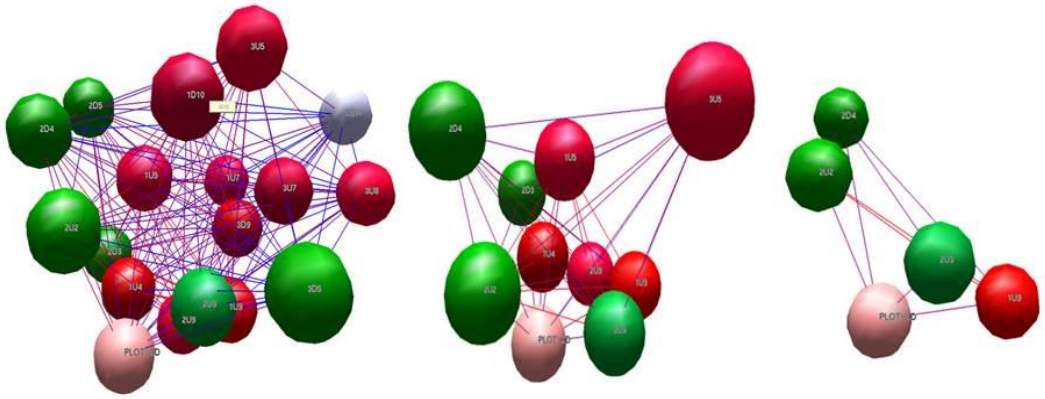
Spearman's ρ≥ 0.3. P-value≤0.03

Fig. 23 Gene network for percentage of plants with common smut(SMUT).



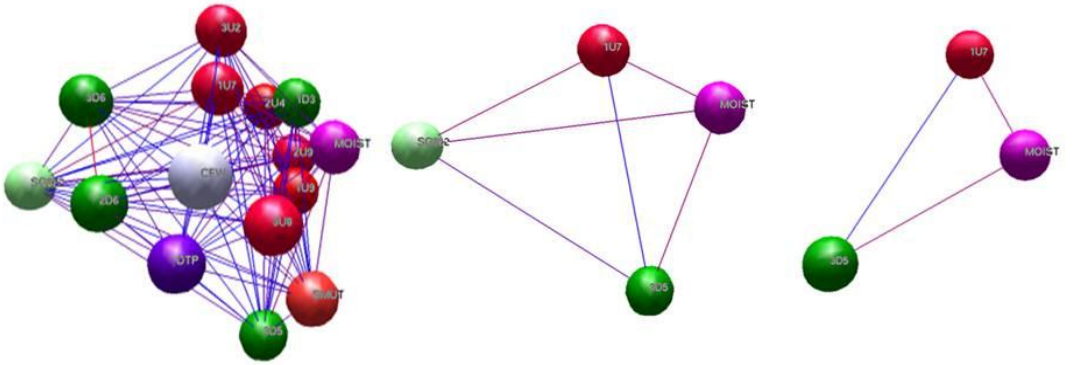
A. Spearman's ρ ≥ 0.3. P-value≤0.03 **B. Spearman's ρ ≥ 0.4. P-value≤0.004**

Fig. 24 Gene network for percentage of molded kernels (MOLD).



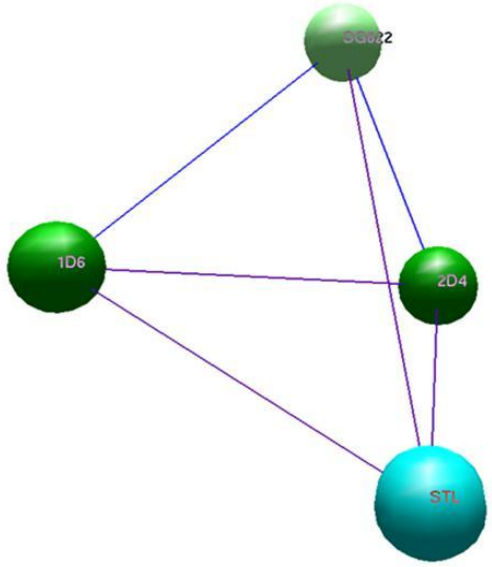
A. Spearman's ρ≥ 0.3 P-value≤0.03 **B. Spearman's ρ≥ 0.4 P-value≤0.004** **C. Spearman's ρ ≥ 0.5 P-value≤0.0001**

Fig. 25 Gene network for plot grain yield"(PLOTYLD).



A. Spearman's ρ≥ 0.3 P-value≤0.03 **B. Spearman's ρ≥ 0.4 P-value≤0.004** **C. Spearman's ρ ≥ 0.5 P-value≤0.0001**

Fig. 26 Gene network for grain moisture at harvest"(MOIST).



Spearman's ρ> 0.3. P-value<0.003

Fig.27 Gene network with trait "Stalk Lodging"(STL).

Table 1 Summary of plant materials.

SAMPLE	TEST	Plot	ENO	Hybrid	GEN
22	WX9C-NDFI	148	1	ARN0902 x B110	F1
33	WX9C-NDFI	225	1	ARN0902 x B110	F1
45	TAES-line-NDFI	139	2	ARN0902	Inbred
60	TAES-line-NDFI	259	2	ARN0902	Inbred
11	WX9C-NDFI	117	4	ARN0902 x B5C2A	F1
31	WX9C-NDFI	222	4	ARN0902 x B5C2A	F1
46	TAES-line-NDFI	129	4	B5C2A	Inbred
61	TAES-line-NDFI	224	4	B5C2A	Inbred
9	WX9C-NDFI	115	5	ARN0902 x CUBA-1	F1
37	WX9C-NDFI	233	5	ARN0902 x CUBA-1	F1
13	WX9C-NDFI	122	6	ARN0902 x S2B73	F1
40	WX9C-NDFI	236	6	ARN0902 x S2B73	F1
1	WX9C-NDFI	101	7	ARN0902 x S2B73BC	F1
43	WX9C-NDFI	249	7	ARN0902 x S2B73BC	F1
47	TAES-line-NDFI	152	8	B110	Inbred
62	TAES-line-NDFI	214	8	B110	Inbred
7	WX9C-NDFI	112	9	BR-1 x B110	F1
30	WX9C-NDFI	221	9	BR-1 x B110	F1
48	TAES-line-NDFI	133	10	B73	Inbred
63	TAES-line-NDFI	235	10	B73	Inbred
49	TAES-line-NDFI	109	11	BR-1	Inbred
64	TAES-line-NDFI	220	11	BR-1	Inbred
18	WX9C-NDFI	132	14	BR-1-B-1 x C273A632	F1
23	WX9C-NDFI	201	14	BR-1x C273A632	F1
12	WX9C-NDFI	119	15	CUBA-1 x B110	F1
36	WX9C-NDFI	232	15	CUBA-1 x B110	F1
50	TAES-line-NDFI	117	15	C3A632-1A	Inbred
65	TAES-line-NDFI	232	15	C3A632-1A	Inbred
16	WX9C-NDFI	127	16	CUBA-1 x BR-1	F1
39	WX9C-NDFI	235	16	CUBA-1 x BR-1	F1
20	WX9C-NDFI	140	18	CUBA-1-B x DK-5-B	F1
35	WX9C-NDFI	230	18	CUBA-1 x DK-5-B	F1
10	WX9C-NDFI	116	20	CUBA-1 x Tx205	F1
34	WX9C-NDFI	226	20	CUBA-1 x Tx205	F1
14	WX9C-NDFI	123	22	DK-5 x B110	F1
28	WX9C-NDFI	217	22	DK-5 x B110	F1
2	WX9C-NDFI	103	24	DK-5 x C3B4	F1
32	WX9C-NDFI	223	24	DK-5 x C3B4	F1
8	WX9C-NDFI	113	25	DK-5 x LH200	F1
25	WX9C-NDFI	208	25	DK-5 x LH200	F1
5	WX9C-NDFI	110	26	DK-7 x B110	F1
27	WX9C-NDFI	216	26	DK-7 x B110	F1
15	WX9C-NDFI	124	27	DK-7 x B5C2A	F1
24	WX9C-NDFI	207	27	DK-7 x B5C2A	F1
51	TAES-line-NDFI	108	27	C273A632	Inbred
66	TAES-line-NDFI	237	27	C273A632	Inbred
21	WX9C-NDFI	147	28	DK-7 x C3B4	F1
38	WX9C-NDFI	234	28	DK-7 x C3B4	F1
4	WX9C-NDFI	108	29	DK-7 x C273A632-1A	F1
29	WX9C-NDFI	219	29	DK-7 x C273A632	F1
52	TAES-line-NDFI	168	29	C3B4	Inbred
67	TAES-line-NDFI	269	29	C3B4	Inbred
17	WX9C-NDFI	129	30	DK-7 x CUBA-1	F1
41	WX9C-NDFI	239	30	DK-7 x CUBA-1	F1
53	TAES-line-NDFI	158	30	CUBA-1	Inbred
68	TAES-line-NDFI	249	30	CUBA-1	Inbred
19	WX9C-NDFI	135	34	DK-7 x S2B73-1	F1
42	WX9C-NDFI	246	34	DK-7 x S2B73-1	F1
54	TAES-line-NDFI	103	34	DK-5	Inbred
69	TAES-line-NDFI	268	34	DK-5	Inbred
6	WX9C-NDFI	111	35	DK-7 x S2B73BC	F1
44	WX9C-NDFI	250	35	DK-7 x S2B73BC	F1
55	TAES-line-NDFI	107	35	DK-7	Inbred
70	TAES-line-NDFI	264	35	DK-7	Inbred
3	WX9C-NDFI	106	45	Tx205 x B110	F1
26	WX9C-NDFI	215	45	Tx205 x B110	F1
56	TAES-line-NDFI	116	46	S2B73	Inbred
71	TAES-line-NDFI	242	46	S2B73	Inbred

Table 1 (Continued)

SAMPLE	TEST	Plot	ENO	Hybrid	GEN
57	TAES-line-NDFI	130	47	S2B73BC	Inbred
72	TAES-line-NDFI	238	47	S2B73BC	Inbred
58	TAES-line-NDFI	151	50	Tx205	Inbred
73	TAES-line-NDFI	217	50	Tx205	Inbred
59	TAES-line-NDFI	105	55	Mo17	Inbred
74	TAES-line-NDFI	216	55	Mo17	Inbred
75	TAES-line-NDFI	371		B73	Inbred
76	TAES-line-NDFI	372		B73 x Mo17	F1
77	TAES-line-NDFI	373		Mo17	Inbred

Table 2 Annotations of genes selected from microarray

ID	gene name	gene label	LogRatio(A/B)	Annotation
MZ00005311	BM382762	1U1	3.324513197	NA
MZ00029551	TC242036	1U2	3.310107111	Zinc-finger protein 1
MZ00012977	CF637688	1U3	3.143782188	apetala2 domain-containing CBF1-like protein
MZ00054057	CF003411	1U4	3.092484325	NA
MZ00052749	BU050513	1U5	2.854918216	NA
MZ00036806	TC221107	1U6	2.841573537	OSJNBa0006B20.16
MZ00056066	TC244413	1U7	2.636032637	NA
MZ00019602	TC233126	1U8	2.58899208	hypothetical protein
MZ00031007	TC243909	1U9	2.528637686	NA
MZ00032181	TC230425	1U10	2.521912244	hypothetical protein
MZ00028889	TC226175	1D1	-3.1622237	OSJNBa0043A12.27
MZ00054188	CF018868	1D2	-1.868682984	NA
MZ00005201	BM379902	1D3	-1.703658013	NA
MZ00015068	TC221657	1D4	-1.692564499	OSJNBb0072M01.18
MZ00024389	TC235186	1D5	-1.667725864	OSJNBa0033G05.15
MZ00029301	TC226390	1D6	-1.670719463	beta-expansin
MZ00013609	TC233770	1D7	-1.591093823	Sucrose-UDP glucosyltransferase 2
MZ00014414	TC235349	1D8	-1.579135915	probable DNA replication licensing factor
MZ00041533	TC235336	1D9	-1.553239644	proliferating cell nuclear antigen
MZ00013463	TC233301	1D10	-1.544707633	contains EST AU164600(R0675) hypothetical protein
MZ00014257	TC235099	2D1	-1.127936193	glucose-1-phosphate adenylyltransferase
MZ00024711	TC220652	2D2	-1.095889716	Probable xyloglucan endotransglucosylase/hydrolase precursor
MZ00041102	TC233774	2D4	-3.818024384	chlorophyll a/b binding protein
MZ00041101	TC233772	2D3	-3.00590284	Chlorophyll a-b binding protein 48, chloroplast precursor
MZ00041103	TC233775	2D5	-3.783839982	Chlorophyll a-b binding protein 48, chloroplast precursor
MZ00015122	TC236479	2U1	1.342996863	putative auxin-repressed protein
MZ00026661	TC222723	2U2	2.023220415	putative NAC-domain protein
MZ00039626	CF014726	2U3	3.248460337	drought-induced hydrophobic protein
MZ00041427	TC220166	2U4	1.158570181	Catalase isozyme 3
MZ00051351	AZM4_79393	2U5	1.378980807	O-methyltransferase ZRP4
MZ00012489	CF625162	2U6	2.482446368	C2H2 type zinc finger transcription factor ZFP37
MZ00029284	TC226636	2U7	2.480264936	At1g10030
MZ00041276	TC234364	2U8	2.471063369	chitinase
MZ00028536	TC240980	2U9	2.442006944	contains ESTs AU083541(S10187),AU070312(S10187)
MZ00036523	BG355384	2U10	2.509304175	NA
MZ00014395	TC220576	2D6	-1.543245218	nucleasome/chromatin assembly factor D protein NFD106
MZ00035785	TC233731	2D7	-1.540034954	beta-glucosidase aggregating factor precursor
MZ00013522	TC218817	2D8	-1.525776746	putative 60S ribosomal protein L38

Table 2 (Continued)

ID	gene name	gene label	LogRatio(A/B)	Annotation
MZ00026784	TC238324	2D9	-1.520517421	contains EST C27797(C52876) unknown protein
MZ00023383	TC233124	2D10	-1.514158501	histone H3
MZ00036799	BI396057	3D9	-3.337027168	NA
MZ00019676	TC227702	3D8	-1.716514162	NA
MZ00037878	CB617078	3D7	-1.438419302	putative diphosphate-fructose-6-phosphate 1-phosphotransferase alpha chain
MZ00013345	TC218442	3D6	-1.326543195	histone H4
MZ00000289	AI001317	3D5	-1.302015625	putative enoyl-ACP reductase
MZ00040034	CF046173	3D4	-1.282462502	NA
MZ00023421	TC233520	3D3	-1.204130095	minichromosomal maintenance factor
MZ00029497	TC242108	3D2	1.063815967	At1g69640/F24J1.22
MZ00010865	CF036161	3D1	1.125626851	putative serine/threonine protein kinase
MZ00016874	TC238176	3U10	1.208074043	putative branched chain alpha-keto acid dehydrogenase E2 subunit
MZ00020151	TC228583	3U9	1.214328651	hypothetical protein F16F4.11
MZ00024521	TC235295	3U8	1.503981954	putative RING protein
MZ00005284	BM382023	3U7	1.511345142	NA
MZ00044565	TC244294	3U6	1.539406871	unknown protein
MZ00026778	TC223933	3U5	1.683480937	putative protein
MZ00017456	TC239586	3U4	1.708021097	protein kinase homolog
MZ00041198	TC220077	3U3	1.861868807	NA
MZ00026392	TC222530	3U2	1.92044857	Bax inhibitor-1 (BI-1)
MZ00036717	BG841531	3U1	1.924561225	ABA- and ripening-inducible-like protein

Table 3 Primer sequence of reference gene

Gene name	Go Number	Forward Primer	Reverse Primer
Elongation factor 1-alpha (EF1-A),	U76259	TTCTCAGTATCCTCCTCTTGG	CAGCCTTCGTCACCTTGG
β-Actin gene(ACT1)	GQ339773	CCTGAGGTTCTATTCCAGCCATCC	GGAGCCACCACTGAGGACAAC
High mobility group protein(HMG)	X58282.1	GGGCAGGAAAGGGAAGGC	CCTCCATGAACACGAAGAAAGC
Glyceraldehyde-3-phosphate dehydrogenase(GAPDH)	EU953063	CACGGCCACTGGAAGCA	TCCTCAGGGTTCCTGATGCC

Table 4 Primer of genes studied

Gene ID	Forward Primer	Reverse Primer
1U1-MZ00005311	ATATAACAACCATAGCAAAGCAC	TTCCAAACCCGTACTTGAATAC
1U2-MZ00029551	TACCGAGTGCGAGAAAAGATG	AATGAGAACCAAGTGTGTATGAAC
1U3-MZ00012977	ATTGTGTTGGCAAAAGGAACC	GTTGGGCACATCTTCTTCTG
1U4-MZ00054057	CATCTCACGCTCACGCTCTG	CAATCGTCTGTTCTTATGCTGCTC
1U5-MZ00052749	CGTTCGTTTAAGATATTGGTCAGC	TGTTGAAAGCAGCAGCCTATG
1U6-MZ00036806	GGAGGCGGTGGCGGCTAC	CGGCGGCGATGGAGTTGAGG
1U7-MZ00056066	CGGTGTACTATGTGTCTTGTTG	GTGACATGCCGAGCGTAG
1U8-MZ00019602	GTTTTCACTTTTGATAATGGGTTC	CCCTAGTGCTACAAAGATTTACAG
1U9-MZ00031007	GGACAAGGACGGCAGGATC	CGAACGAACAAAGAATTTCTGAAC
1U10-MZ00032181	GTCCGAACTGTACCACCTG	CCACTGTTCCACGATCCATTC
1D1-MZ00028889	CCACCTTATGTCGTCCTTCAC	CTGGTTAGGCTGTAATTCATTGC
1D2-MZ00054188	TTGCGGTAATGCTGTTAG	CAATTATGAAATGGACAAGTC
1D3-MZ00005201	ATTAGATGGAAGTAGAACACACAC	TCAGCACTAGAATGTTTGGAC
1D4-MZ00015068	AGGTGAGCATCCAATTCGTGAG	TCGGAGGGCTTGGTGACC
1D5-MZ00024389	TAGCATAGCAGCAACCAGCAATG	GAGCGGCGACGAGGAGAG
1D6-MZ00029301	GGTGTCAGCAGTAATTTAAGCAAG	CGTCGTCTTATGCCTGTCTATC
1D7-MZ00013609	ACCCTAACTCCGACCTCTACTGG	TGAACTGGCACGAGAAGTGGTAG
1D8-MZ00014414	CACGCCTTAGTTGGTCTCTG	ACCAGTTACACGGCACAC
1D9-MZ00041533	CTCTTCCTTCCAGCCGTTG	ACAAACCCAGTAAATCAGTAATCC
1D10-MZ00013463	GCCGAAGGTTATGTTTTGAAGTTC	ACATAGGCACAAGGTAATCCG
2D1-MZ00014257	TGGAGTAGTGCGAACAGTG	GCGTCCGAGTTTATTATTTACAAG
2D2-MZ00024711	TATTGTTGTGTGTCAGGTTCG	TTCGGTGTGAGGAGATACG
2D3-MZ00041101	AGAACCTCGCCGACCACATC	TGCCCTCGCTCACTTGCC
2D4-MZ00041102	GCTTACGCCACCAACTTC	GCTCCCTTCACAAATACTGG
2D5-MZ00041103	GACCCCGTCAACAACAAC	GCTCCCTTCACAAATACTGG
2U1-MZ00015122	GGGATGAGTTGCGGGTTTAC	GCATAGCACCTTATTGATACAGC
2U2-MZ00026661	CTGGTCGGGGTTACAAGTC	GCAGGAAGTGAAACTAAAGAATG
2U3-MZ00039626	CTGCCTCCTCCTCACCTTCCTC	GTGATGGCGTAGATGGCGTAGATG
2U4-MZ00041427	TGCCCGTGTCCGCCTTTG	TTTGCTGCTGGTGCCTTAATTTG
2U5-MZ00051351	TGCTGGACCTCCCGAACG	GCAGAACCCACTTGAGGAAGAC
2U6-MZ00012489	CCAGCGGCGGCATTGATTAG	GAAGCAAAGCAAGCAAGCAACC
2U7-MZ00029284	CCAGCGGACGGAGAAAGAG	GAAGAGGCAGGACCACACG
2U8-MZ00041276	GGAACAACCCCGCCCAGATG	CGCCGTCGCCTAGCAAGTG
2U9-MZ00028536	CATGCGTAGAGAATTGAGATGTC	GTAGGCAGCACTTCCATTGAG
2U10-MZ00036523	ATCTATGTTGTTGTGGACTTGTGG	CCCGTATAGTAAGGTGCGATGG
2D6-MZ00014395	TGCCACTAAGAAAGCCAAGACAG	ACCATCGTCATCCTCGTCCTC
2D7-MZ00035785	TTCCCGAGACTTCCCTGACATC	CTCCATCGTATCATCCAGCAACC
2D8-MZ00013522	GCTAATTCGCCCATCCATCG	TTTCTAATGCCAATGAACACTGAG
2D9-MZ00026784	TCGTACCATGTCTGATCCCATC	AACAGTCATAACAGTACACCGTAG

Table 4 (Continued)

Gene ID	Forward Primer	Reverse Primer
2D10-MZ00023383	CGTGTTGTCCTCACTGAATGG	ATTATGTTGGCATGTTATGGTTCC
3U1-MZ00036717	GGGGAGGGACGACGATTC	CCCAAGAAGCCCAAGGAG
3U2-MZ00026392	AGAGGAAGAAGAGGTCGTGAGC	GAACAAAGGTGCCGAACAAAGG
3U3-MZ00041198	CACCCTTCAGTTATCTTTGGACAG	GCCTACCTACATAATCAGTGATGC
3U4-MZ00017456	ATGGTGTTGTCAAGATTCAGG	CCGCCCGTTTGTTTCATC
3U5-MZ00026778	GCACACTAGAAGAACACCTTTTAC	CGTCGGCACCTTCGCTATG
3U6-MZ00044565	TCAGAGTTTGGTTCAGGCATTG	CTACAAGTGGGTGCGAGAGG
3U7-MZ00005284	CACCACTTCCGTTTGCTGAAAG	GCGACACTGCCAACACTGC
3U8-MZ00024521	AGCCACGAAGAGACCTGAC	CGTAGCAAGTCTGATGTCCAC
3U9-M00020151	CAGAGTTCATGCCACCGAGAAAC	CTCAGAGGATGGACCACCAGATC
3U10-MZ00016874	AATGAATGGAAGAGCCTCGTG	CTCCGTTTCAGGTCGTAGTATG
3D1-MZ00010865	CCGAACCGCCTCCTTTATAG	GCCTACTATTGCTGTTGATACTC
3D2-MZ00029497	GCTGTGGCTTCCTGGTAATATC	ATGTAAGTTCCGAGAATCTTGTCC
3D3-M00023421	TTCAGGGACAACGGCTTC	TATCATCACTCTTCATCTCAGG
3D4-MZ00040034	TCGTCTTGGTCCGTCGTC	GAAACTGAATACATCCCTGAATG
3D5-MZ00000289	CACGGGTTTCTAGGACTGATTG	ACACTCTTCCGATATTGCTCTTAC
3D6-MZ00013345	TCCTCATCTGGTGGCGAACTG	GCGACGATATAGAGAACGGACTAC
3D7-MZ00037878	CGACAGAGCAGGTCAATGGTG	GAGCAGCGTCAGTGTTGGAAG
3D8-MZ00019676	CTTCTCAGTAGTTGGCTTCAC	TGCTGTTGCGGTTCATTATC
3D9-MZ00036799	CTCCATGTTCGGCTTCTTC	CCAGGCGTTGTTGTTGAC

Table 5 Summary of 13 phenotypic traits.

Genotype	DTP	PHT	EHT	STL	RTL	SG822	Husk	CEW	EL	SMUT	Mold	PLOTYLD	MOIST
ARN0902 x B110	70	225	86	0	0	2.5	2.5	6.6	18.8	0	10	14.66	17.4
ARN0902 x B110	73	236	105	0	0	2	3	6	16	0	4	18.11	17.9
ARN0902 x B5C2A	73	228	94	0	0	2.5	3	6.8	19.6	0	3	16.18	16.3
ARN0902 x B5C2A	73	222	96	0	0	2	3	5.2	18.8	0	10	20.13	16.0
ARN0902 x CUBA-1	72	225	104	0	0	2.5	2	6.6	19.2	0	3	17.58	15.2
ARN0902 x CUBA-1	71	224	108	0	0	3	2.5	5.6	19.4	5	3	18.58	15.1
ARN0902 x S2B73	69	208	95	0	0	4	2.5	4	17.6	0	2	13.43	14.6
ARN0902 x S2B73	69	220	85	0	0	3	2	3.4	18	0	2	16.71	15.0
ARN0902 x S2B73BC	70	223	91	0	0	2.5	3	6	18.8	0	5	15.09	15.8
ARN0902 x S2B73BC	72	226	100	0	0	2	2	6.4	19.2	5	2	17.59	15.4
BR-1 x B110	74	256	110	2	0	2.5	3	5	21	0	3	17.96	16.5
BR-1 x B110	74	258	120	0	2	2.5	3	5.8	21.2	0	3	22.06	15.4
BR-1-B-1 x C273A632	73	235	113	2	1	2.8	3	5	18.6	5	1	12.56	15.0
BR-1-B-1 x C273A632	74	252	103	1	0	3	4	5	18.6	0	1	18.79	14.6
CUBA-1 x B110	74	236	95	2	0	4	3	4.4	15.4	0	10	16.68	16.1
CUBA-1 x B110	73	260	119	0	3	3.5	3	5	17.2	0	1	18.14	15.3
CUBA-1 x BR-1	75	260	105	0	0	2.5	3	6.4	20	0	1	12.92	14.8
CUBA-1 x BR-1	75	255	119	0	0	2	3.5	2.4	24	0	1	12.56	15.7
CUBA-1 x Tx205	73	225	95	0	0	2.5	3.5	7	17.2	5	5	16.49	15.4
CUBA-1 x Tx205	74	247	90	0	0	2.5	3	3.8	18.4	10	1	14.29	15.5

Table 5 (Continued)

Genotype	DTP	PHT	EHT	STL	RTL	SG822	Husk	CEW	EL	SMUT	Mold	PLOTYLD	MOIST
CUBA-1-B x DK-5-B	71	230	98	0	0	3	2.5	4.8	17.8	0	1	20.65	16.0
CUBA-1-B x DK-5-B	70	254	101	1	0	3	3	4.6	18.4	10	1	17.80	15.6
DK-5 x B110	72	212	93	2	0	3.5	3	4	17.6	10	3	15.65	16.4
DK-5 x B110	71	238	116	7	0	3	3	5	19.6	0	2	17.83	16.6
DK-5 x C3B4	71	210	76	0	0	3	4	4	18	5	10	13.31	17.4
DK-5 x C3B4	71	209	67	0	0	2.5	3.5	2.2	17.2	0	1	15.94	17.9
DK-7 x B110	74	243	94	4	8	3	3	3.6	17.8	10	20	16.56	17.6
DK-7 x B110	73	251	111	0	0	3	3.5	4.6	18.8	10	3	20.29	17.8
DK-7 x B5C2A	76	246	116	0	0	1.5	3	5.8	18.2	20	3	17.65	19.4
DK-7 x B5C2A	76	229	106	0	0	1.5	3.5	3.4	18	20	5	14.65	21.0
DK-7 x C273A632-1A	73	232	99	0	0	3	3	4.8	17.4	0	6	12.65	14.8
DK-7 x C273A632-1A	74	234	108	0	0	3.5	3.5	4.4	17.4	10	5	10.80	15.5
DK-7 x C3B4	74	225	88	2	1	2.5	3	2	16.8	5	1	14.34	18.1
DK-7 x C3B4	75	240	88	1	0	2.5	3.5	2.2	18	20	1	13.81	19.0
DK-7 x CUBA-1	73	260	110	0	0	2.5	3	4.2	19.8	20	2	15.48	17.6
DK-7 x CUBA-1	73	249	92	0	0	1.5	3	3.4	18.8	5	5	17.57	18.7
DK-7 x S2B73-1	72	258	102	1	0	3	3.5	3.8	17.8	5	2	16.38	15.3
DK-7 x S2B73-1	71	263	103	0	0	2.5	3.5	3.2	18	10	1	13.93	15.5
DK-7 x S2B73BC	74	245	99	0	0	2.5	4	4.2	17.2	10	8	16.44	17.4
DK-7 x S2B73BC	75	259	111	0	0	2	3.5	3.8	18	0	2	15.74	18.7
Tx205 x B110	73	248	101	2	0	3	3	4.4	17.4	15	8	14.40	15.6
Tx205 x B110	73	258	110	0	0	3.5	3	5	18.4	0	5	14.76	14.7

Table 6 Correlation of 39 genes among inbreds at level Spearman's ρ≥0.6

Variable	by Variable	Spearman ρ	Prob>|ρ|
1U9	1U4	0.7809	<.0001
1D9	1U9	0.6505	<.0001
1D10	1U9	0.6643	<.0001
1D10	1D9	0.8808	<.0001
2U1	1D9	0.6505	<.0001
2U2	1U2	0.6448	<.0001
2U2	1U4	0.7132	<.0001
2U3	1U4	0.7839	<.0001
2U3	1U9	0.632	<.0001
2U3	2U2	0.625	<.0001
2U4	1U4	0.7159	<.0001
2U9	1D9	0.651	<.0001
2U9	1D10	0.6423	<.0001
2U9	2U1	0.6753	<.0001
2U9	2U4	0.6389	<.0001
2D1	1D9	0.7775	<.0001
2D1	1D10	0.7562	<.0001
2D1	2U9	0.6605	<.0001
2D3	1D9	0.7768	<.0001
2D3	1D10	0.765	<.0001
2D3	2D1	0.8378	<.0001
2D4	1D9	0.7604	<.0001
2D4	1D10	0.6981	<.0001
2D4	2D1	0.8498	<.0001
2D4	2D3	0.919	<.0001
2D5	1D9	0.6817	<.0001
2D5	1D10	0.6531	<.0001
2D5	2D1	0.7578	<.0001
2D5	2D3	0.8694	<.0001
2D5	2D4	0.8814	<.0001
2D6	1D9	0.844	<.0001
2D6	1D10	0.7635	<.0001
2D6	2D1	0.6823	<.0001
2D6	2D3	0.7648	<.0001
2D6	2D4	0.6795	<.0001
2D6	2D5	0.7695	<.0001
2D7	1D9	0.6362	<.0001
2D7	1D10	0.6433	<.0001
2D9	1D3	0.6	<.0001
2D9	1D9	0.8248	<.0001
2D9	1D10	0.7291	<.0001
2D9	2D1	0.7241	<.0001
2D9	2D3	0.7801	<.0001
2D9	2D4	0.7097	<.0001
2D9	2D5	0.8002	<.0001
2D9	2D6	0.9448	<.0001
2D10	2D3	0.6237	<.0001
2D10	2D5	0.6952	<.0001
2D10	2D6	0.6375	<.0001
2D10	2D9	0.666	<.0001

Table 6 (Continued)

Variable	by Variable	Spearman ρ	Prob>|ρ|
3U5	2D5	0.6327	<.0001
3U5	3U4	0.744	<.0001
3U7	3U5	0.6071	<.0001
3U9	2D10	0.6004	<.0001
3U10	1D9	0.6589	<.0001
3U10	1D10	0.7012	<.0001
3U10	2D3	0.6612	<.0001
3U10	2D4	0.6161	<.0001
3U10	2D5	0.7005	<.0001
3U10	2D6	0.7055	<.0001
3U10	2D9	0.6959	<.0001
3U10	3U4	0.68	<.0001
3U10	3U5	0.6494	<.0001
3D1	2D5	0.6178	<.0001
3D1	2D6	0.6207	<.0001
3D1	2D9	0.6196	<.0001
3D1	3U4	0.699	<.0001
3D1	3U10	0.7828	<.0001
3D2	2D3	0.6045	<.0001
3D2	2D5	0.699	<.0001
3D2	2D6	0.6367	<.0001
3D2	2D9	0.6276	<.0001
3D2	3U4	0.6457	<.0001
3D2	3U5	0.6716	<.0001
3D2	3U9	0.6425	<.0001
3D2	3U10	0.9009	<.0001
3D2	3D1	0.7538	<.0001
3D3	2D7	0.6292	<.0001
3D4	2D6	0.6163	<.0001
3D4	2D7	0.6247	<.0001
3D4	2D9	0.667	<.0001
3D4	3U7	0.6264	<.0001
3D9	1D10	0.6245	<.0001
3D9	2D1	0.6077	<.0001
3D9	2D3	0.7291	<.0001
3D9	2D5	0.68	<.0001
3D9	2D6	0.7252	<.0001
3D9	2D9	0.7623	<.0001
3D9	2D10	0.601	<.0001
3D9	3U4	0.6523	<.0001
3D9	3U5	0.6659	<.0001
3D9	3U9	0.6529	<.0001
3D9	3U10	0.75	<.0001
3D9	3D1	0.759	<.0001
3D9	3D2	0.7218	<.0001

Table 7 Correlation of 39 genes among hybrids at level Spearman's ρ≥0.6

Variable	by Variable	Spearman ρ	Prob>|ρ|
1U4	1U2	0.6603	<.0001
1U5	1U2	0.6609	<.0001
1U5	1U4	0.9107	<.0001
1U7	1U5	0.6989	<.0001
1U9	1U4	0.8737	<.0001
1U9	1U5	0.9275	<.0001
1U9	1U7	0.7153	<.0001
1D3	1U5	0.6419	<.0001
1D10	1U5	0.6072	<.0001
1D10	1D9	0.663	<.0001
2U2	1U4	0.84	<.0001
2U2	1U5	0.7743	<.0001
2U2	1U9	0.771	<.0001
2U3	1U2	0.6084	<.0001
2U3	1U4	0.8357	<.0001
2U3	1U5	0.8593	<.0001
2U3	1U9	0.8494	<.0001
2U3	2U2	0.8523	<.0001
2U4	1U5	0.6709	<.0001
2U4	1U7	0.7041	<.0001
2U4	1U9	0.6358	<.0001
2U4	2U3	0.6541	<.0001
2U9	1U5	0.6487	<.0001
2U9	1U9	0.672	<.0001
2U9	2U2	0.6052	<.0001
2U9	2U3	0.6895	<.0001
2U9	2U4	0.7031	<.0001
2D1	1D9	0.8054	<.0001
2D1	1D10	0.6833	<.0001
2D3	1U4	0.6803	<.0001
2D3	1U5	0.7851	<.0001
2D3	1U9	0.7044	<.0001
2D3	1D3	0.6367	<.0001
2D3	2U3	0.6592	<.0001
2D3	2U4	0.6007	<.0001
2D4	2U9	0.6623	<.0001
2D4	2D3	0.6161	<.0001
2D5	1D9	0.6302	<.0001
2D5	2U4	0.6239	<.0001
2D5	2D1	0.6541	<.0001
2D5	2D3	0.6794	<.0001
2D5	2D4	0.6349	<.0001
2D6	2D1	0.688	<.0001
2D9	1D9	0.6313	<.0001
2D9	2D1	0.7821	<.0001
2D9	2D6	0.8072	<.0001
2D10	1D6	0.6045	<.0001
2D10	1D9	0.6775	<.0001

Table 7 (Continued)

Variable	by Variable	Spearman ρ	Prob>|ρ|
2D10	1D10	0.6154	<.0001
2D10	2D1	0.7851	<.0001
2D10	2D6	0.6678	<.0001
2D10	2D9	0.7817	<.0001
3U5	2U1	0.7415	<.0001
3U5	2D6	0.6232	<.0001
3U7	1D9	0.6545	<.0001
3U7	2D1	0.6371	<.0001
3U7	2D9	0.6815	<.0001
3U7	2D10	0.7523	<.0001
3U8	3U5	0.6095	<.0001
3U10	1D9	0.6413	<.0001
3U10	2D1	0.7901	<.0001
3U10	2D5	0.6221	<.0001
3U10	2D9	0.6989	<.0001
3U10	2D10	0.6867	<.0001
3U10	3U7	0.6837	<.0001
3D1	2D1	0.6048	<.0001
3D1	2D6	0.6145	<.0001
3D1	2D9	0.6126	<.0001
3D1	3U10	0.7499	<.0001
3D2	1D9	0.6285	<.0001
3D2	2D1	0.6939	<.0001
3D2	2D5	0.6591	<.0001
3D2	2D9	0.6049	<.0001
3D2	2D10	0.7315	<.0001
3D2	3U7	0.6891	<.0001
3D2	3U10	0.7868	<.0001
3D3	2D1	0.6215	<.0001
3D3	2D6	0.6856	<.0001
3D3	2D9	0.6905	<.0001
3D3	2D10	0.6592	<.0001
3D4	1U5	0.6197	<.0001
3D4	2U1	0.6062	<.0001
3D6	2D1	0.7108	<.0001
3D6	2D6	0.8036	<.0001
3D6	2D9	0.7524	<.0001
3D6	2D10	0.6798	<.0001
3D6	3U10	0.6148	<.0001
3D6	3D3	0.7253	<.0001
3D9	1U5	0.6763	<.0001
3D9	1U7	0.6573	<.0001
3D9	1U9	0.6299	<.0001
3D9	1D3	0.6582	<.0001
3D9	2U4	0.6625	<.0001
3D9	2D3	0.6612	<.0001
3D9	2D5	0.7189	<.0001
3D9	3D2	0.6312	<.0001
3D9	3D4	0.6622	<.0001

Table 8 Correlation of 39 genes among eMPH at level Spearman's ρ≥0.6

Variable	by Variable	Spearman ρ	Prob>|ρ|
1U4	1U2	0.6733	<.0001
1U5	1U2	0.7093	<.0001
1U5	1U4	0.8733	<.0001
1U9	1U2	0.6615	<.0001
1U9	1U4	0.8238	<.0001
1U9	1U5	0.905	<.0001
1D3	1U7	0.6572	<.0001
1D4	1D3	0.665	<.0001
1D7	1D3	0.6417	<.0001
1D7	1D4	0.7054	<.0001
1D9	1D4	0.6208	<.0001
1D10	1D4	0.6095	<.0001
1D10	1D9	0.7075	<.0001
2U1	1D9	0.7731	<.0001
2U1	1D10	0.6275	<.0001
2U2	1U4	0.8467	<.0001
2U2	1U5	0.7704	<.0001
2U2	1U9	0.76	<.0001
2U3	1U4	0.7959	<.0001
2U3	1U5	0.778	<.0001
2U3	1U9	0.757	<.0001
2U3	2U2	0.7705	<.0001
2U9	1U7	0.6338	<.0001
2U9	1D3	0.611	<.0001
2U9	1D4	0.6328	<.0001
2U9	1D9	0.7039	<.0001
2U9	1D10	0.7114	<.0001
2U9	2U1	0.6468	<.0001
2U9	2U4	0.604	<.0001
2D1	1D4	0.6939	<.0001
2D1	1D9	0.7654	<.0001
2D1	1D10	0.783	<.0001
2D1	2U1	0.6412	<.0001
2D1	2U9	0.831	<.0001
2D3	1U7	0.6679	<.0001
2D3	1D3	0.7636	<.0001
2D3	1D4	0.675	<.0001
2D3	1D10	0.6161	<.0001

Table 8 (Continued)

Variable	by Variable	Spearman ρ	Prob>|ρ|
2D3	2U9	0.7557	<.0001
2D3	2D1	0.7734	<.0001
2D4	1D3	0.6208	<.0001
2D4	1D4	0.6954	<.0001
2D4	1D10	0.6068	<.0001
2D4	2U9	0.7292	<.0001
2D4	2D1	0.8048	<.0001
2D4	2D3	0.882	<.0001
2D5	1U7	0.6344	<.0001
2D5	1D3	0.7032	<.0001
2D5	1D4	0.6393	<.0001
2D5	1D9	0.6164	<.0001
2D5	1D10	0.6721	<.0001
2D5	2U9	0.7163	<.0001
2D5	2D1	0.7996	<.0001
2D5	2D3	0.8789	<.0001
2D5	2D4	0.7848	<.0001
2D6	1D9	0.7832	<.0001
2D6	1D10	0.7448	<.0001
2D6	2U1	0.6294	<.0001
2D6	2U9	0.7588	<.0001
2D6	2D1	0.8899	<.0001
2D6	2D3	0.6375	<.0001
2D6	2D4	0.6179	<.0001
2D6	2D5	0.7119	<.0001
2D7	2D1	0.6161	<.0001
2D7	2D6	0.687	<.0001
2D9	1D9	0.7426	<.0001
2D9	1D10	0.7115	<.0001
2D9	2U9	0.7248	<.0001
2D9	2D1	0.8111	<.0001
2D9	2D5	0.6515	<.0001
2D9	2D6	0.9165	<.0001
2D9	2D7	0.662	<.0001
2D10	2D1	0.6662	<.0001
2D10	2D5	0.6225	<.0001
2D10	2D6	0.672	<.0001
2D10	2D9	0.6782	<.0001
3U4	2D5	0.621	<.0001
3U4	2D10	0.624	<.0001

Table 8 (Continued)

Variable	by Variable	Spearman ρ	Prob>|ρ|
3U7	1D10	0.6266	<.0001
3U7	2D9	0.6603	<.0001
3U8	1U5	0.6122	<.0001
3U10	1D9	0.6397	<.0001
3U10	1D10	0.647	<.0001
3U10	2U1	0.6111	<.0001
3U10	2U9	0.6221	<.0001
3U10	2D1	0.7689	<.0001
3U10	2D5	0.669	<.0001
3U10	2D6	0.8214	<.0001
3U10	2D9	0.7981	<.0001
3U10	2D10	0.6663	<.0001
3U10	3U7	0.6154	<.0001
3D1	1D9	0.6664	<.0001
3D1	1D10	0.6509	<.0001
3D1	2U1	0.6497	<.0001
3D1	2D1	0.7397	<.0001
3D1	2D6	0.759	<.0001
3D1	2D9	0.7398	<.0001
3D1	3U10	0.8094	<.0001
3D2	1D10	0.6297	<.0001
3D2	2D1	0.7664	<.0001
3D2	2D3	0.6054	<.0001
3D2	2D5	0.7209	<.0001
3D2	2D6	0.7764	<.0001
3D2	2D7	0.6787	<.0001
3D2	2D9	0.7401	<.0001
3D2	2D10	0.6194	<.0001
3D2	3U4	0.6555	<.0001
3D2	3U7	0.6501	<.0001
3D2	3U10	0.883	<.0001
3D2	3D1	0.7319	<.0001
3D3	2D6	0.6601	<.0001
3D3	2D7	0.6214	<.0001
3D3	2D9	0.7013	<.0001
3D3	3U7	0.6743	<.0001
3D3	3U10	0.6834	<.0001
3D3	3D2	0.7391	<.0001
3D4	1D3	0.6228	<.0001
3D4	2U4	0.6774	<.0001

Table 8 (Continued)

Variable	by Variable	Spearman ρ	Prob>|ρ|
3D4	2D3	0.6073	<.0001
3D4	2D5	0.6858	<.0001
3D6	1D10	0.663	<.0001
3D6	2U9	0.635	<.0001
3D6	2D1	0.8531	<.0001
3D6	2D3	0.6351	<.0001
3D6	2D5	0.7378	<.0001
3D6	2D6	0.8079	<.0001
3D6	2D9	0.7646	<.0001
3D6	3U10	0.8645	<.0001
3D6	3D1	0.7998	<.0001
3D6	3D2	0.8675	<.0001
3D6	3D3	0.6547	<.0001
3D9	1U7	0.6021	<.0001
3D9	1D3	0.6447	<.0001
3D9	1D10	0.6674	<.0001
3D9	2U9	0.6189	<.0001
3D9	2D1	0.6677	<.0001
3D9	2D3	0.732	<.0001
3D9	2D5	0.7843	<.0001
3D9	3U7	0.6793	<.0001
3D9	3U10	0.6736	<.0001
3D9	3D1	0.6054	<.0001
3D9	3D2	0.6911	<.0001
3D9	3D4	0.6563	<.0001
3D9	3D6	0.7249	<.0001

Table 9 Correlation between genes and traits at level Spearman's ρ⩾0.6

Traits	Gene name	QTL location of gene	correlated genes	Spearman ρ	Prob>|ρ|	gene annotation
DTP	MZ00005201	q1000k1/qcobd1	1D3	0.4273	0.0048	NA
DTP	MZ00015068	qdpoll1_QTL_days_to_pollen_1	1D4	0.4913	0.001	OSJNBb0072M01.18
DTP	MZ00041533	qearht1_QTL_ear_height_1/qearl1_QTL_ear_length_1	1D9	0.5696	<.0001	proliferating cell nuclear antigen
DTP	MZ00014257	qearl1_QTL_ear_length_1	2D1	0.3778	0.0136	glucose-1-phosphate adenylyltransferaselarge subunit
DTP	MZ00014395	qepp1_QTL_ears_per_plant_1/qearl1_QTL_ear_length_1	2D6	0.4828	0.0012	nucleasome/chromatin assembly factor D protein
DTP	MZ00026784	qeard1_QTL_ear_diameter_1/qearht1_QTL_ear_height_1	2D9	0.3446	0.0254	contains EST C27797(C52876)
DTP	MZ00023383	qtest1_QTL_test_weight_1/vgt1_vegetative_to_generative_transition1_	2D10	0.3485	0.0237	histone H3 {Zea mays;}
DTP	MZ00020151	qearht1_QTL_ear_height_1	3U9	0.539	0.0002	hypothetical protein F16F4.
DTP	MZ00010865	qkrow1_QTL_kernel_row_number_1/qkrowl1_QTL_kernel_row_length_1/qplht10_QTL_plant_height_10	3D1	0.4293	0.0046	putative serine/threonine protein kinase
DTP	MZ00029497		3D2	0.3264	0.0349	At1g69640/F24J1.22
PHT	MZ00015068	qdpoll1_QTL_days_to_pollen_1	1D4	0.513	0.0005	OSJNBb0072M01.18
PHT	MZ00041533	qearht1_QTL_ear_height_1/qearl1_QTL_ear_length_1	1D9	0.5871	<.0001	proliferating cell nuclear antigen
PHT	MZ00015122	qearl1_QTL_ear_length_1	2U1	0.3498	0.0231	putative auxin-repressed protein
PHT	MZ00014395	qepp1_QTL_ears_per_plant_1/qearl1_QTL_ear_length_1	2D6	0.3264	0.0349	nucleasome/chromatin assembly factor D protein

Table 9 (Continued)

Traits	Gene name	QTL location of gene	correlated genes	Spearman ρ	Prob>|ρ|	gene annotation
PHT	MZ00037878	qearht1_QTL_ear_height_1/qearl1_QTL_ear_length_1	3D7	0.3137	0.0431	putative diphosphate-fructose-6-phosphate 1-phosphotransferase alpha chain
PHT			DTP	0.4576	0.0023	
EHT	MZ00029551	q1000k1	1U2	-0.323	0.037	Zinc-finger protein 1
EHT	MZ00015068	qdpoll1_QTL_days_to_pollen_1	1D4	0.3553	0.021	OSJNBb0072M01.18
EHT	MZ00041533	qearht1_QTL_ear_height_1/qearl1_QTL_ear_length_1	1D9	0.37	0.0159	proliferating cell nuclear antigen
EHT	MZ00017456	qearl1_QTL_ear_length_1	3U4	0.3455	0.025	protein kinase homolog
EHT	MZ00037878	qearht1_QTL_ear_height_1/qearl1_QTL_ear_length_1	3D7	0.3982	0.009	putative diphosphate-fructose-6-phosphate 1-phosphotransferase alpha chain {
EHT			DTP	0.3518	0.0223	
EHT			PHT	0.6177	<.0001	
STL	MZ00029301		1D6	-0.3297	0.033	beta-expansin
SG822	MZ00056066		1U7	0.5414	0.0002	NA
SG822	MZ00013463	qdpoll1_QTL_days_to_pollen_1/qeard1_QTL_ear_diameter_1/qearht1_QTL_ear_height_1	1D10	-0.3295	0.0331	contains EST AU164600(R0675) hypothetical protein
SG822	MZ00029497		3D2	-0.321	0.0382	At1g69640/F24J1.22
SG822	MZ00023421		3D3	-0.4377	0.0037	minichromosomal maintenance factor
SG822	MZ00000289		3D5	-0.3173	0.0406	minichromosomal maintenance factor
SG822	MZ00013345	qcobd1/qdpoll1_QTL_days_to_pollen_1	3D6	-0.374	0.0147	minichromosomal maintenance factor
SG822			DTP	-0.3627	0.0183	
SG822			STL	0.3508	0.0227	
Husk	MZ00054057	q1000k1/q300k1/qcobd1	1U4	0.391	0.0105	NA

Table 9 (Continued)

Traits	Gene name	QTL location of gene	correlated genes	Spearman ρ	Prob>|ρ|	gene annotation
Husk	MZ00052749	qkrow1_QTL_kernel_row_number_1/qkrowl1_QTL_kernel_row_length_1	1U5	0.3247	0.0359	NA
Husk	MZ00031007	qdpoll1_QTL_days_to_pollen_1	1U9	0.3179	0.0402	NA
Husk	MZ00005201	q1000k1/qcobd1	1D3	0.4255	0.005	NA
Husk	MZ00015068	qdpoll1_QTL_days_to_pollen_1	1D4	0.4826	0.0012	OSJNBb0072M01.18
Husk	MZ00041533	qearht1_QTL_ear_height_1/qearl1_QTL_ear_length_1	1D9	0.3504	0.0229	proliferating cell nuclear antigen
Husk	MZ00026661	qearl1_QTL_ear_length_1	2U2	0.3368	0.0292	putative NAC-domain protein
Husk	MZ00039626	qtest1_QTL_test_weight_1	2U3	0.3179	0.0402	drought-induced hydrophobic protein
Husk	MZ00020151	qearht1_QTL_ear_height_1	3U9	0.3363	0.0295	hypothetical protein F16F4.11
Husk			DTP	0.4249	0.005	
CEW	MZ00054057	q1000k1/q300k1/qcobd1	1U4	-0.3571	0.0202	NA
CEW	MZ00052749	qkrow1_QTL_kernel_row_number_1/qkrowl1_QTL_kernel_row_length_1	1U5	-0.3671	0.0168	NA
CEW	MZ00031007	qdpoll1_QTL_days_to_pollen_1	1U9	-0.3874	0.0113	NA
CEW	MZ00029301		1D6	0.4014	0.0084	beta-expansin
CEW	MZ00026661		2U2	-0.3995	0.0088	minichromosomal maintenance factor
CEW	MZ00039626	qtest1_QTL_test_weight_1	2U3	-0.3787	0.0134	drought-induced hydrophobic protein
CEW	MZ00026778	qgrwt1_QTL_grain_weight_1	3U5	-0.3777	0.0149	putative protein
CEW			Husk	-0.3894	0.0108	
EL	MZ00016874		3U10	0.4227	0.0053	putative branched chain alpha-keto acid dehydrogenase E2 subunit
EL			EHT	0.3549	0.0211	NA

Table 9 (Continued)

Traits	Gene name	QTL location of gene	correlated genes	Spearman ρ	Prob>|ρ|	gene annotation
EL			SG822	-0.3286	0.0336	
EL			CEW	0.3439	0.0257	
SMUT	MZ00056066		1U7	-0.3887	0.011	NA
SMUT	MZ00020151	qearht1_QTL_ear_height_1	3U9	0.3514	0.0225	hypothetical protein F16F4.11
SMUT			CEW	-0.3634	0.018	Bax inhibitor-1 (BI-1)
Mold	MZ00029551	q1000k1	1U2	0.3187	0.0397	Zinc-finger protein 1
Mold	MZ00005201	q1000k1/qcobd1	1D3	0.3079	0.0473	NA
Mold	MZ00026392	qgrwt1_QTL_grain_weight_1	3U2	-0.4594	0.0022	Bax inhibitor-1 (BI-1)
Mold	MZ00037878	qearht1_QTL_ear_height_1/qearl1_QTL_ear_length_1	3D7	0.4133	0.0065	putative diphosphate-fructose-6-phosphate 1-phosphotransferase alpha chain
E10W			EL	0.5126	0.0005	
PLOTYLD	MZ00054057	q1000k1/q300k1/qcobd1	1U4	-0.4893	0.001	NA
PLOTYLD	MZ00052749	qkrow1_QTL_kernel_row_number_1/qkrowl1_QTL_kernel_row_length_1	1U5	-0.4641	0.002	NA
PLOTYLD	MZ00056066		1U7	-0.3447	0.0254	NA
PLOTYLD	MZ00031007	qdpoll1_QTL_days_to_pollen_1	1U9	-0.5803	<.0001	NA
PLOTYLD	MZ00013463	qdpoll1_QTL_days_to_pollen_1/qeard1_QTL_ear_diameter_1/qearht1_QTL_ear_height_1	1D10	-0.3645	0.0176	contains EST AU164600(R0675) hypothetical protein
PLOTYLD	MZ00026661		2U2	-0.548	0.0002	putative NAC-domain protein
PLOTYLD	MZ00039626	qtest1_QTL_test_weight_1	2U3	-0.539	0.0002	drought-induced hydrophobic protein
PLOTYLD	MZ00028536	qearht1_QTL_ear_height_1	2U9	-0.4194	0.0057	unknown protein

Table 9 (Continued)

Traits	Gene name	QTL location of gene	correlated genes	Spearman ρ	Prob>|ρ|	gene annotation
PLOTYLD	MZ00041101		2D3	-0.4452	0.0031	Chlorophyll a-b binding protein 48, chloroplast precursor
PLOTYLD	MZ00041102		2D4	-0.5112	0.0009	chlorophyll a/b binding protein
PLOTYLD	MZ00041103		2D5	-0.3172	0.0407	Chlorophyll a-b binding protein 48, chloroplast precursor
PLOTYLD	MZ00026778	qgrwt1_QTL_grain_weight_1	3U5	-0.4429	0.0037	putative protein
PLOTYLD	MZ00005284		3U7	-0.3907	0.0105	NA
PLOTYLD	MZ00024521		3U8	-0.3155	0.0418	putative RING protein
PLOTYLD	MZ00000289		3D5	0.3114	0.0447	putative enoyl-ACP reductase
PLOTYLD	MZ00036799		3D9	-0.3069	0.0481	NA
PLOTYLD			CEW	0.3756	0.0142	
MOIST	MZ00056066		1U7	-0.5158	0.0005	NA
MOIST	MZ00031007	qdpoll1_QTL_days_to_pollen_1	1U9	-0.3074	0.0476	NA
MOIST	MZ00005201	q1000k1/qcobd1	1D3	0.3139	0.0429	NA
MOIST	MZ00041427		2U4	-0.3539	0.0215	Catalase isozyme 3
MOIST	MZ00028536	qearht1_QTL_ear_height_1	2U9	-0.3892	0.0108	unknown protein
MOIST	MZ00014395	qepp1_QTL_ears_per_plant_1/qearl1_QTL_ear_length_1	2D6	0.3244	0.0361	nucleasome/chromatin assembly factor D protein NFD106
MOIST	MZ00026392	qgrwt1_QTL_grain_weight_1	3U2	-0.3538	0.0215	Bax inhibitor-1 (BI-1)

Table 9 (Continued)

Traits	Gene name	QTL location of gene	correlated genes	Spearman ρ	Prob>|ρ|	gene annotation
MOIST	MZ00024521		3U8	-0.3752	0.0144	putative RING protein
MOIST	MZ00000289		3D5	0.5344	0.0003	putative enoyl-ACP reductase
MOIST	MZ00013345	qcobd1/qdpoll1_QTL_days_to_pollen_1	3D6	0.3185	0.0398	histone H4 - Arabidopsis thaliana
MOIST			SG822	-0.4871	0.0011	
MOIST			CEW	-0.3352	0.03	
MOIST			SMUT	0.3719	0.0153	

Table 10 Summery of number of genes and traits correlated to 12 phenotypic traits.

Trait	No. of Genes Correlated to Trait			Correlated Traits(Spearman's ρ > 0.3)
	Spearman's ρ > 0.3	Spearman's ρ > 0.4	Spearman's ρ > 0.5	
Stalk Lodging (STL)	2	0	0	SG822
Days to Pollen Shedding (DTP)	10	6	2	PHT, EHT, HUSK MOIST
Plant Height (PHT)	5	2	2	EH, DTP
Ear Height (EHT)	5	0	0	PHT,STL,EL,SG822, DTP
Stay Green Rate （SG822）	6	2	1	STL, EL, DTP, PLOTYLD, MOIST
Husk Coverage Rate (HUSK)	9	2	0	DTP CEW
Corn Earworm Feeding Damage (CEW)	7	1	0	PLOTYLD, MOIST
Ear Length (EL)	1	1	0	CEW, EHT,SG822, PLOTYLD
Percentage of Plants with Common Smut (SMUT)	2	0	0	MOIST, CEW
Percentage of Molded Kernels (MOLD)	4	2	0	
Plot Grain Yield (PLOTYLD)	16	9	4	CEW, EL, SG822
Grain Moisture at Harvest(MOIST)	10	2	2	CEW, DTP, SMUT SG822

Table 11 Comparison of results in this study with those of Thiemann et al. 2010

Gene ID	Gene name	Traits associated with genes identified in this study	Traits associated with genes identified by Thiemann et al(2010).
MZ00056066	1U7	SMUT, SG822, PLOTYLD, MOIST	HP_GDMC
MZ00015068	1D4	PHT, EHT, HUSK, DTP	HP_GDMC
MZ00028536	2U9	PLOTYLD, MOIST, DTP	MPH_GY
MZ00028383	2D10	DTP	HP_GDMC, HP_GY, MPH_GY
MZ00040034	3D4		HP_GDMC, MPH_GY
MZ00023421	3D3	HUSK, SG822	MPH_GY
MZ00016874	3U10	EL	HP_GDMC
MZ00026392	3U2	MOLD, MOIST	HP_GY, MPH_GY

VITA

Name:	Hui Zhi
Address:	Department of Soil and Crop Sciences, Texas A&M University, 370 Olsen Blvd., 2474 TAMU, College Station, TX 77843-2474, USA
Email Address:	zxapply@neo.tamu.edu
Education:	B.S., Biotechnology, Beijing Forestry University, 2007 M.S., Plant Breeding, Texas A&M University, 2010